REPORT RESUMES

ED 013 242 SP 001 275

DEVELOPMENT OF A PERFORMANCE TEST OF TEACHING PROFICIENCY. FINAL REPORT.

BY- POPHAM, W. JAMES
CALIFORNIA UNIV., LOS ANGELES
REPORT NUMBER BR-5-0566 PUB DATE AUG 67
CONTRACT OEC-6-10-254
EDRS PRICE MF-$0.75 HC-$6.84 172P.

DESCRIPTORS- COLLEGE STUDENTS, *HIGH SCHOOL STUDENTS, *LAY TEACHERS, *SOCIAL SCIENCES, STUDENT ATTITUDES, *STUDENT BEHAVIOR, TABLES (DATA), TEACHER ATTITUDES, *TEACHER QUALIFICATIONS, TESTS, WONDERLIC PERSONNEL TEST, SOCIAL SCIENCE PRETEST, SOCIAL SCIENCE POSTTEST,

IN CONTRAST TO RATINGS OF OBSERVED TEACHING BEHAVIOR, THE CRITERION OF PUPIL GROWTH WAS USED IN VALIDATING A PERFORMANCE TEST OF TEACHING PROFICIENCY. USING AN INSTRUCTIONAL UNIT ON "SOCIAL SCIENCE RESEARCH METHODS," IT WAS HYPOTHESIZED THAT PUPIL GAINS WOULD BE GREATER AMONG THOSE TAUGHT BY EXPERIENCED THAN BY INEXPERIENCED TEACHERS. PARTICIPANTS WERE GIVEN IN ADVANCE A STATEMENT OF OBJECTIVES IN OPERATIONAL TERMS (TERMINAL BEHAVIORS EXPECTED OF STUDENTS) AND A COLLECTION OF POSSIBLE LEARNING ACTIVITIES AND WERE INSTRUCTED TO TEACH FOR THESE OBJECTIVES. IN THE FIRST OF TWO VALIDATION STUDIES, 6 EXPERIENCED SOCIAL SCIENCE STUDENT TEACHERS AND 6 HOUSEWIVES (FORMER SOCIAL SCIENCE MAJORS) TAUGHT PAID VOLUNTEERS FOR A 6-HOUR PERIOD ON ONE DAY. IN THE SECOND STUDY, 13 REGULARLY CREDENTIALED TEACHERS AND 13 COLLEGE STUDENTS TAUGHT SUMMER SESSION STUDENTS FOR A 4-HOUR PERIOD ON ONE DAY. IN THE FIRST STUDY, STUDENTS TOOK THE WONDERLIC PERSONNEL TEST, A 33-ITEM PRETEST AND A 68-ITEM POSTTEST, AND COMPLETED A QUESTIONNAIRE MEASURING REACTIONS TO THE INSTRUCTION. TEACHERS ALSO COMPLETED AN ASSESSMENT QUESTIONNAIRE. IN THE SECOND STUDY, THE POSTTEST AND THE TWO QUESTIONNAIRES WERE USED. IN NO INSTANCE WERE ANY SIGNIFICANT DIFFERENCES FOUND BETWEEN THE TWO GROUPS OF TEACHERS OR BETWEEN THE STUDENTS TAUGHT BY THESE TEACHERS. "TEACHERS' LACK OF EXPERIENCE IN ACHIEVING PRESET BEHAVIOR CHANGES IN LEARNERS" IS OFFERED IN PARTIAL EXPLANATION OF THE NO-DIFFERENCE FINDINGS. (AF)

01275

ED013242

OCT 4 1967

R 600

FINAL REPORT
Project No. 5-0566-2-12-1
Contract No. OE-6-10-254

DEVELOPMENT OF A PERFORMANCE TEST OF TEACHING PROFICIENCY

August 1967

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

DEVELOPMENT OF A PERFORMANCE TEST
OF TEACHING PROFICIENCY

Project No. 5-0566-2-12-1
Contract No. OE-6-10-254

W. James Popham

August 1967

The research reported herein was performed pursuant to a contract with the Office of Education, United States Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

University of California
Los Angeles, California

ACKNOWLEDGEMENTS

Considerable appreciation is extended to the teachers and administrators of the San Diego City Schools who participated in the conduct of this project. Quite literally, without their efforts the investigation could not have been consummated. In particular, the initiative of Harmon H. Kurtz, Director of Administrative Research, and H. David Fish, Specialist in Social Studies, was critically important in expediting the conduct of the project.

The investigation itself depended largely on the performance test materials which were developed under the supervision of Eva L. Baker. Special recognition should be given for her direction of this activity. Her principal assistants in this endeavor were Adrianne Bank and Lily Cooper, both of whom made significant contributions.

W.J.P.

CONTENTS

Page

INTRODUCTION 1

METHOD 6

    Validation Study I 8
    Validation Study II 12

RESULTS 14

DISCUSSION 18

CONCLUSIONS 20

SUMMARY 21

APPENDICES A-1

INTRODUCTION

There is no need to document the need for reliable measures of teaching proficiency. In almost unlimited instances educators could profit enormously if valid indices of teaching effectiveness were available. It is partly because of this tremendous need that one despairs when surveying the advances made by researchers who have worked in the area of teacher-competence assessment during the past 60 years. The distressing truth is that we have not moved very far forward in our efforts to develop reliable measures of teacher effectiveness.

A review of teacher effectiveness research since the turn of the century finds a considerable number of studies, but none which might be classified as "breakthrough." Along with the periodic reviews of the field, such as those by Morsh and Wilder[1] and, more recently, Barr[2] and and Ryans,[3] we have witnessed a number of theoretical and empirical reports by individuals such as Cogan, Flanders, Gage, Medley, Smith, Wilk, Edson, and Withall,[4] to mention but a few.

The Dilemma. There seems to be general agreement that the ultimate criterion of teacher competence should be pupil growth, and we have witnessed innumerable efforts to predict such growth through the use of systematic observations and ratings of teaching behavior. Countless hours have been devoted to the careful observation and categorization of what teachers do in the classroom. In recent years the sophistication with which observers have attacked classroom problems is almost frightening. Observation schemes and explanatory constructs have become so multidimensional that the classroom observer may soon be forced to choose among several hundred descriptive dimensions every few milliseconds. Ratings scales, too, have become more and more complicated and we find greater efforts devoted to the analysis of increasingly smaller aspects of the teacher's performance.

---

[1] Joseph E. Morsh and Eleanor Wilder, "Identifying the Effective Instructor: A Review of the Quantitative Studies, 1900-1952," *Research Bulletin AFPTRC-TR-54-44*, Lackland Air Force Base, Texas, 1954.

[2] A. S. Barr, issue editor, "Wisconsin Studies of the Measurement and Prediction of Teacher Effectiveness: A Summary of Investigations," *Journal of Experimental Education*, Vol. 30, September, 1961, pp. 5-156.

[3] David G. Ryans, "Assessment of Teacher Behavior and Instruction," *Review of Educational Research*, Vol. 33, No. 4, October, 1963, pp. 415-441.

[4] All of these researchers contributed articles to the Symposium on Classroom Behavior of Teachers, edited by Harry F. Silberman for the Vol. 14, No. 3, September, 1963, issue of *The Journal of Teacher Education*.

In essence, most commonly used measures of teacher effectiveness are attempts to mirror in some way the teacher's probable success in promoting learner achievement. Such procedures as ratings and check lists are generally conceded to be estimates, albeit very gross ones, of the teacher's ability to promote learner growth. The problem with such measures is that learner growth may be directed toward extremely different ends. Different teachers may have markedly divergent goals in mind for their students. Therefore, when an evaluator such as a principal or supervisor observes a teacher in action and gives him a rating, it is usually the evaluator who is implicitly imposing his own conception of what the teacher's goals should be when he makes his rating. The truth is that the teacher's goals may be at considerable variance with those of the evaluator.

A second, more common situation occurs when a rater is concerned with only the instructional means which the teacher employs, without any explicit consideration of the ends the teacher is trying to achieve. In such instances, the evaluator may still rate the teacher according to the evaluator's personal standards regarding what form classroom activities should take. But even here the evaluator's ideas regarding desirable classroom procedures will be inextricably tied to his notions, however cloudy, of what the outcomes of instruction should be. Therefore, it is not surprising that ratings, check lists, and other comparable approaches have yielded very unsatisfactory results with respect to the assessment of teacher competence. McNeil[5] has eloquently enumerated the the problems encountered when evaluators tend to impose their own value systems on teachers through the use of supposedly "neutral" measurement devices.

Faced with the complexities resulting from divergent instructional objectives, those researchers studying classroom teaching procedures make a very critical mistake when they attempt to ferret out supposedly "superior" instructional procedures which could be used with equal efficacy by different teachers. Most experienced researchers in this field now recognize that the quality of learning which transpires in a given instructional situation is a function of particular instructional procedures employed by a particular instructor for particular students with particular goals in mind. In other words, the instructional means may vary considerably from one teacher to another and yet both could accomplish identical ends with equal success.

Why not turn directly to measures of pupil growth in an effort to avoid this problem? Once again we encounter teachers' committments to dissimilar objectives which make it difficult to apply evaluation devices such as standardized achievement tests. The comprehensive nature of standardized measures often obscures the very important differential emphases made by teachers during the instructional process.

---

[5] McNeil, John D., "Antidote to a School Scandal." Educational Forum 31:69-77; November 1966.

Rationale. This report describes results of using a different approach[6] to this assessment dilemma. At UCLA we have attempted to circumvent the "good teaching procedures" question and use only pupil growth as a criterion. For the present we are willing to overlook the question of how teachers achieve pupil gain. We recognize that because of fantastic personal variation among teachers, extremely different procedures may work, even when "work" is defined in terms of pupils' achieving precisely the same objectives. We certainly believe that researchers should continue their efforts to discover which instructional techniques have a high probability of achieving instructional ends. But for the moment, we shall simply side-step the question of what those techniques are.

Our assumption is that the teacher who is the better achiever of given instructional goals will, other factors being relatively equal, be the better achiever of his own goals. We therefore are developing tests of teaching performance wherein the teacher is given sets of explicit instructional objectives, asked to teach specifically to them, and has his instructional competence assessed in terms of his ability to produce the pupil behavior changes described by those objectives. A set of performance tests of instructor competence is currently being developed at UCLA under provision of two U.S.O.E. contracts. Two of the tests are in the field of vocational education and one is in the social sciences.

This report describes the development and two field tests of the social science performance test. These tests consist of (1) a set of operational instructional objectives stated in terms of specific pupil behaviors, (2) a collection of possible learning activities which a teacher may wish to employ, and (3) pre- and post-tests, not seen or administered by the teacher, which adhere closely to the operational objectives. The objectives and possible activities are given to the teacher well in advance of instruction and he is told to prepare plans for a specified period of teaching. Members of the project staff administer the pre- and post-tests to the teacher's pupils and their progress toward the objectives serves as an index of the teacher's proficiency. By stipulating identical objectives to be achieved, but permitting teacher diversity in the means used to accomplish these ends, a method of evaluating teaching performance without restricting individualistic teaching style is provided.

Since the initial report[7] of this line of research, several other investigators have described comparable approaches to the question of teacher competence assessment. For example, Moffett[8] has reported an attempt to evaluate student teachers by the degree to which they can obtain prespecified instructional objectives. Medley[9] has also described a series of goal-oriented teaching exercises.

---

[6] Briefly described in an "Of Special Significance" report, The Journal of Teacher Education 17:250-51; Summer 1966.

[7] Popham, W. J., and Baker, E. L., "A Performance Test of Teaching Effectiveness." Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, Illinois, February 17-19, 1966.

Theoretical Problems. There are several theoretical problems associated with this particular approach to teacher competence assessment. One of the most important concerns the validation of the performance tests which have been developed. Our plan has been to attempt a validation wherein the performance of nonteachers (housewives or college students, for example) is pitted against that of experienced teachers. The validation hypothesis predicts that the experienced teachers will secure better pupil achievement than will the nonteachers. This particular hypothesis, of course, is an extremely gross test in that we wish to secure a marked contrast between (1) those that have never taught, and (2) those who have taught for some time.

One of the problems rests on whether the experienced teachers will by truly "good" teachers for this kind of contrast. To what extent will mere experience prove advantageous to regularly credentialed instructors in achieving prespecified behavior changes in learners? Indeed, are "experienced" teachers really *experienced* in modifying pupil behavior according to previously established objectives?

Other problems are associated with the sensitivity of the measuring instruments we have developed. Can the pre- and post-tests based on the operational objectives be made sufficiently reliable and discriminating to serve our purposes? A very early field trial[10] of one of the performance tests revealed scant differences between instructors who were teaching classes at very different levels of advancement in a junior college vocational program. If we cannot secure performance differences between such markedly different classes, surely our efforts are to be in vain. Moreover, it is unlikely that tests which only yield statistical significance will be of any practical value in teacher competence assessment. We need to develop ways of discriminating between teachers in a much more dramatic fashion.

Another difficulty is associated with the attitudes of teachers toward the particular objectives selected. We will certainly make every effort to select instructional objectives which are approved by the teachers in the field, but undoubtedly there will be differences in the way these objectives are received. Perhaps we can control, in part, the difference in the way teachers view our objectives by administering preinstruction attitudinal measures to instructors regarding the objectives. But we may also find that some teachers differ in the degree to which they will accept imposed objectives in preference to those that they might choose themselves.

---

[8] Moffett, George M., "Use of Instructional Objectives in the Supervision of Student Teachers." Unpublished doctoral dissertation. University of California, Los Angeles, 1966.

[9] Medley, Donald M., and Hill, Russell, *Brief for Project GOTE*. Mimeographed, Educational Testing Service, October 26, 1966.

[10] Popham, W. J., and Baker, E. L., *op. cit.*

Practical Problems. There are also very practical problems associated with these tests. One such problem is securing cooperation on the part of the public schools. We have encountered a fair amount of difficulty in securing participation by California school districts. Even though it is generally conceded that research on this topic could be of considerable value to school districts, their policies ofter preclude such investigations. Frankly, many districts are reluctant to assist in the development of teacher proficiency tests based on pupil achievement. Perhaps some administrators will fear intrusion on their teacher-evaluation domain. Some teachers may feel genuinely threatened by the prospect of having their competence assessed in terms of pupil achievement. Until the performance tests have been demonstrated to be valid and helpful measures of at least this one aspect of teaching competence, we can anticipate some opposition to the tryout of our instruments.

Another problem is related to the procedures associated with the validation test of these performance measures. Will the nonteachers' lack of familiarity with the pupils they are teaching prove to be the major deterrent to their successful accomplishment of the objectives? Must we setup a situation in which the nonteachers have an opportunity to work previously with the classes in order to allow them equal familiarity with the pupils? Or, on the other hand, should we setup these situations so that we stack the deck as much as possible in favor of the experienced teachers? Since we wish to demonstrate that at the grossest level these tests can discriminate between two extreme groups of instructors, perhaps we serve our purpose better if we do let the experienced teachers work with their own classes while the inexperienced teachers instruct unfamiliar pupils.

Related to this problem is the associated one that the nonteacher may attend more carefully to the given objectives because this is, perhaps, the most obvious way of structuring his thinking regarding the instructional task. The experienced teacher undoubtedly has a number of predilections regarding instructional pursuits, even though this subject matter unit may be unique in his school's curriculum. We may find that experienced teachers actually suffer far more from distractors than the inexperienced teachers and, therefore, will tend to perform less effectively when performance is assessed exclusively by their ability to accomplish such prespecified instructional goals.

METHOD

This section will describe the preparation of the performance test material and the conduct of two attempts to validate the instrument.

The Production of the Unit. The more tedious and time-consuming phases of this project involved the selection of a topic, statement of objectives, assembly of resource materials, and construction of test items. Any topic chosen for the unit had to meet several requirements. First, it should take 10 hours or less to teach. This limited time per period would, we hoped, permit us to secure more cooperation from public school personnel who might be willing to devote two weeks of normal class time to our field trials but would be loath to give longer periods to the research. Second, it should not be currently taught in the schools. This was to reduce the likelihood of previous student exposure to the material. Third, the topic should require no specific set of student entry behaviors dependent upon previous instruction. For example, we could not presume that any special knowledge (such as historical facts) were in the repertoire of the learners. Fourth, the topic should be able to be inserted logically at any point in the curriculum. For instance, a unit on preliterate societies would meet the other two criteria above, but its inclusion in an American history course would be absurd. This requirement of logic also pointed to a general social science topic, equally relevant in a geography or civics class and not based exclusively on one particular subject field. A last requirement for the unit topic was that it be acceptable to teachers, and they would feel it important enough to employ. The reason for the latter requirement is obvious.

An area which seemed to satisfy handily the five criteria above was one dealing with the "interrelationships among the social sciences." Sets of social science teachers and supervisors were interviewed at the junior and senior high schools levels and were generally enthusiastic about such a topic. Seven major social sciences were ch[illegible] and research assistants began preparing material for the resource unit. Lists of distinctive characteristics and shared characteristics for each social science were to be compiled. Tentative objectives called for the students to identify (1) characteristics distinctive to given social sciences, (2) characteristics shared by different social science fields, and (3) problems peculiar to particular social sciences. Despite much work and the encouragement of local teachers, this topic was eventually abandoned, for while only little agreement was found regarding the distinctive characteristics of different fields, there was complete chaos when one tried to classify social problems as being the proper domain of one or another social science.

The topic which was finally chosen six months later concerned research methodology in the social sciences and was entitled "Social Science Research Methods." The objectives of the unit call for the student to perform many tasks familiar to the research worker. Students are asked to discriminate between causal and descriptive hypothesis, to identify major advantages and limitations of different data collection procedures, to interpret data in graph form, to define important terms, and to match famous social scientists with descriptions of their work. In addition, when given brief descriptions of research, students should be able to identify (1) basic components of the research, (2) broad classes of error in the design and conduct of the research (for example, sampling problems or lack of authentication), (3) relevant information omitted from research summaries, and (4) the type of analysis procedure employed in a given study.

Resource Unit. The resource unit was organized in the following way: First, a brief introduction and overview of the unit was presented with a list of general objectives. Next, 13 specific objectives were listed, each followed by a sample test item. The next section was titled "Content Guidelines." Included in it was all information necessary for one to answer the test items correctly. This part of the unit was required because during the assembly of resources rather large differences were found to exist in the way experts defined certain terms (for example, the word "experiment"). In view of these areas of confusion, we decided to define operationally some of the terms. For instance, "internal authentication" procedures were defined as applying only to the actual content of any communication and "experiment," whenever used, meant that a condition was manipulated by the researcher.

The "Content Guidelines" section, as all subsequent material, was organized by topic beginning with hypothesis and data collection techniques and ending with analysis procedures. Following this section was a more general set of resources, including 25 summaries of actual research. The unit consisted of 26 pages. A copy of it is found in Appendix A.

Preparation of Tests. A version of these objectives and the sample test items were given to 18 social science graduate students and teachers for their comments. All reported that this was an acceptable topic. Revisions in the objectives were generally made consistent with their suggestions. Their greatest objection was that the test items called for selected rather than constructed responses. This was not changed.

A pool of items was devised for each of the 13 objectives which resulted in a test of 196 responses. On an item sampling basis (each student responding to one quarter of the total number of items) the test was given to 97 junior high school students in the Spring of 1966. Data were analyzed to determine the base line of responses and to identify items which appeared to be ambiguously phrased. The test was again revised, given to 56 more eighth grade students, and eventually an item pool of 125 questions evolved. All questions required selected responses. The difficulty experienced by junior high students in reading and understanding the language of the test led us to use senior high school age students in subsequent projects.

In addition, a sample of five senior high school students and five graduate social science majors were given a copy of the test. The five graduate students were previously given relevant materials from the resource unit and told to answer the questions consistent with the information specified in the unit. Mean score for the "naive" high school students was 70.5. Mean score for the graduate students was 104.8. A rough discrimination score was computed by subtracting the graduate student errors on a given item from the number of errors committed by the high school students. In 90 cases there was no difference (in large part because no one missed the particular items) and in five cases the difference was in the wrong direction. High school students were asked to mark any questions they though were too easy and the graduates were asked to indicate those items which were ambiguous to them. Using all the information obtained as a guide, items were deleted and a 33 item pretest (found in Appendix B) and 68 item post-test (found in Appendix C) were constructed on the basis of the remaining items.

Validation Study I

The initial comparison of teachers and nonteachers was conducted in February, 1967 under conditions more stringently controlled than ordinarily found in school settings and was confined to six consecutive hours of instruction. Two groups, one of six nonteachers and one of six experienced teachers were employed. Each subject taught three or four high school students for a six hour period on one day. The unusually high concentration of instruction followed by immediate testing was designed to aid in the discovery of achievement differences promoted by members of the two groups without the many typical distractions of an actual school setting. Students were pretested and post-tested by members of the research staff.

Sample. The subjects were selected three weeks prior to the day of the study. Members of the experienced-teacher group had the following qualifications: (1) they were social science majors; (2) they had completed at least one quarter of student teaching in which they were judged superior by their supervisors; and (3) they had received a grade of "A" in a pre-service curriculum and instruction class which emphasized the use of behavioral objectives and the careful planning of instruction to achieve these goals. It was believed that teachers who had such qualifications would tend to focus on the precise objectives of the unit, and would choose or devise relevant instructional activities which should lead to the acquisition of the desired behaviors. Of over 185 student teachers who had completed at least one quarter of student teaching, only 12 met all the stated criteria. Six of those agreed to participate.

The comparison group was composed of six nonteachers. These individuals, all housewives, were recruited from the University married student housing facility after an announcement was distributed describing the requirements and offering $25 for participating. (All identical honorarium was given to the experienced teachers.) Requirements were that the participant: (1) should not currently be enrolled in school; (2) should be without formal teaching experience; (3) should have completed at least two years of college; and (4) should have been a social science major.

The requirements of social science major and minimum of two years of college were included so that the content of the unit would be within the individual's experience. The restrictions regarding current enrollment and formal teaching experience were included to assure that the volunteer was not currently a participant in a teaching or learning situation. Six nonteachers were selected at random from those who volunterred. All subjects were mailed a copy of the resource unit, including the list of 13 objectives and sample test items for each. An enclosed letter related that the purpose of the study was to try out a new teaching unit for social studies. They also received a set of directions telling them where to report, that they would have six hours in which to teach, that they would be teaching a group of three or four high school students, and that they should try to teach all of the objectives.

Learners for the instruction were secured from three Los Angeles area high schools. Students were told they would be paid $10 for their services. From 105 volunteers, 50 junior and senior high school students were randomly selected. These students were informed by mail of the location and time they were to report.

Procedure. Prior to the Saturday of the study, students were randomly assigned to the 12 teachers. This was done to control for dissimilarities between groups prior to instruction, although verification of group comparability with respect to ability and entering behavior was planned.

Learners reported at 8:45 in the morning, and the Wonderlic Personnel Test, a 12 minute test of mental ability, was administered. The learners were next allowed 15 minutes to complete the 33 item pretest selected from the item pool of criterion behaviors. Students were then assigned and dispersed to their rooms. In the meantime, the six experienced teachers and six nonteachers had reported to separate assembly rooms and had been directed to their individual classrooms.

At 9:30 a.m. all learners and teachers were in their designated places and each of the 12 teachers commenced his instruction. After a 45 minute lunch break, instruction was resumed and continued until 4:00 p.m., at which time the high school students and the teachers assembled in different rooms. The students were first given the 68 item post-test measuring each of the explicit objectives. They next completed a questionnaire (found in Appendix D) designed to measure their feelings about the content of the unit and the instruction they received.

During this time, the teachers completed a questionnaire which asked for their assessment of the materials, their opinion of objectives, their suggestions for revisions, and their appraisal of the procedures and conditions under which the study was conducted.

Results. The post-test was scored so that a top score of 80 was possible. The mean post-test score for students of the experienced teachers was 59.0, and for the nonteachers was 58.7. The difference was obviously minor. Means and standard deviations for the post-test, pretest, and Wonderlic by teacher are presented in Table 1.

There were no systematic differences between the affective questionnaire responses of pupils taught by teachers and those taught by nonteachers. There were also no differences between the responses of the teachers and nonteachers to the instructor questionnaires.

Table 1. Means and Standard Deviations of Post-Test, Pretest, and Wonderlic Scores by Teacher: Validation Study I

Subjects	n	Post-Test		Pretest		Wonderlic	
		$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
Teachers							
1	4	59.2	2.5	11.2	3.6	21.7	7.4
2	4	58.2	6.9	8.4	2.9	20.2	5.2
3	3	57.6	9.7	12.3	3.8	23.0	8.0
4	4	67.2	5.7	14.6	6.8	24.7	5.6
5	4	51.5	8.5	9.8	0.7	18.5	3.8
6	3	60.0	4.3	15.8	7.0	23.0	5.5
Nonteachers							
1	4	56.5	12.3	9.5	3.6	21.0	7.4
2	3	59.3	9.0	12.4	2.3	24.3	3.5
3	4	63.7	6.5	14.2	4.6	24.2	6.7
4	3	64.0	1.4	14.3	1.5	22.6	1.1
5	4	58 0	12.3	12.3	9.0	26.7	6.8
6	4	6[illegible].7	3.0	8.9	3.0	16.0	1.7

Means and standard deviations for post-test, pretest, and Wonderlic by group are presented in Table 2.

Table 2. Means and Standard Deviations of Post-Test, Pretest, and Wonderlic Scores by Group: Validation Study I

Group	n	Post-Test		Pretest		Wonderlic	
		$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
Teachers	22	59.0	1.89	11.84	1.54	21.59	1.64
Nonteachers	22	58.7	3.07	12.87	1.73	24.24	2.12

In an effort to identify possible equating variables, an intercorrelation matrix was constructed using the following factors: pretest score, Wonderlic score, grade point average, affective questionnaire score, teachers' questionnaire score, and post-test score. The correlation coefficient obtained for the pretest and post-test was .68 and the correlation between the Wonderlic Personnel Test and the post-test was .77, indicating that these two measures might be useful covariates.

An estimate of the internal consistency of the post-test was provided by the Kuder-Richardson (21) formula. The Kuder-Richardson (21) coefficient was .82.

Discussion. Unfortunately, the clarity of retrospect is one of the greatest frustrations in the conduct of educational research. When one has the perverse opportunity to explore alternative explanations for the data, as in the case of this study, the number of competing factors is uncomfortably large. The disappointingly slight differences between effects produced by teachers and nonteachers could be accounted for by any of the following reasons. First, nonteachers were originally chosen as the comparison group because they would probably feel less confident in a classroom environment and therefore, it was supposed, be less effective instructionally. Perhaps this would have been true in a school classroom where the teacher would be forced to content with strange and complex stimuli. But the conditions under which this study was conducted were very controlled and both teacher and nonteacher were in complete command of a few students in a quiet room. There was really little else to occupy the subjects' thoughts but the unit objectives. In addition, from observation, the nonteachers who volunteered were not shy, retiring housewife-types, but rather self-confident, aggressive women who did not at all appear to be intimidated by the prospect of teaching three of four high school students. It is also interesting to note that all subjects, teachers and nonteachers, reported that they had indeed intended to teach all thirteen objectives to their students. And it now seems obvious that nonteachers might even have cleaved more closely to the specifications in the unit, especially when they lacked teaching experience.

The comparatively good performance of the nonteachers as opposed to the experienced teachers might also be explained in relation to the usual experiences of the two groups. For the housewives, the participation in the study was a rather special event, a distinct departure from their normal activities. For the teachers, however, teaching more students on a Saturday was probably a rather drab prospect. It is possible, then, that the results reflect these different perceptions of the task, with the nonteacher exerting extra effort and bringing more enthusiasm to their teaching.

The results can also in part be attributed to the type of learners recruited for the study. Even though high school administrators were told that we wished "ordinary" students to participate, the students we secured and chose by random procedures were, in fact, quite special. Their mean grade point average was 3.1 (A = 4.0), and average Wonderlic score was well above the mean for high school graduates. These students were so bright that the quality of the teaching may have had little facilitating or retarding effect on their achievement. The task as well may not have been difficult enough to challenge the learners. Originally the test had been designed for junior high school students but had been proven too difficult. Perhaps, it was too easy for bright high school juniors and seniors.

Largely as a consequence of these interpretations, the procedure used in the second validation study was altered considerably. A description of that study now follows.

Validation Study II

The second validation study was conducted in the San Diego City Schools during the 1957 summer session. In brief, the following modifications from the initial study were incorporated: (1) The investigation was conducted in a regular school setting. (2) The time allowed for instruction was somewhat reduced. (3) College students, rather than housewives, were used as the nonteachers. (4) Since regular classes were randomly divided into two smaller classes, one taught by the regular teacher and one taught by the nonteacher, no pretest was used.

Sample. The teacher subjects were 13 instructors in regular San Diego City Schools summer session 12th grade government classes. All of the teachers (approximately 20) in such classes were invited[11] to participate in the project. Thirteen agreed. The nonteachers were 13 upper division female college students from San Diego State College. College students were deliberately selected so that the nonteachers would, as the regular teachers, have competing demands (such as study requirements) for their attention and would not be able to make the project their "sole interest" as might have been the case with the housewives in the first study. Requirements for the nonteachers were the following: (1) they had to be social science majors or minors; (2) they must have completed two years of college; (3) they could have no formal teaching experience; and (4) they had taken no college course-work in professional education. Both teachers and nonteachers were given a $25 honorarium for their participation.

Learners for the second validation study were regularly enrolled students in the 12th grade government classes. For the most part, these learners were taking the course for the first time, as opposed to summer school "retakes" by students who had previously failed the course.

Procedure. A one week period at the approximate mid-point of the summer session, July 10-14, was designated as the period during which the project would be carried out. Two weeks prior to that time the resource unit was given to all teachers and nonteachers along with written directions[12] describing the procedure to be followed on the day of their participation in the study.

---

[11] A copy of the invitation letter is included in Appendix F.

[12] These directions are found in Appendix F.

The summer school classes were held from approximately 8:00 a.m. to noon. The period of instruction was one full four hour period. This time reduction, from approximately six hours in the initial study, was made because it was believed to be more consonant with the abilities of the 12th grade students. The day set for the actual teaching was determined at the mutual convenience of the particular teacher and nonteacher involved with each class.

At the close of the class session prior to the day the teaching was to take place, the regular teacher divided his class into two groups by having them "count off by twos" (each student alternately calling out "one" or "two").[13] All of the "ones" were to remain with the regular teacher on the following day. All of the "twos" were to be taught by the nonteachers.

On the day of instruction the nonteacher arrived at the school early and reported to the regular teacher. When the class arrived, the nonteacher took all of the "twos" to a different classroom and commenced the instruction. For legal purposes, a regularly credentialed substitute teacher was also in the nonteacher's room. These substitutes remained unobtrusively at the rear of the classroom and did not interfere with the instructional activities. While the nonteacher was thus engaged, the regular teacher was instructing his half of the class in the regular classroom. Both teachers and nonteachers were told that they could assign homework if they wished.

On the day following instruction all pupils met with their regular teacher who opened a sealed envelope and administered the post-test and student questionnaire at the beginning of the period. Forty minutes were allowed for completion of the test and 10 minutes were given for pupils to finish the questionnaire. During this same time the teacher completed an instructor questionnaire (found in Appendix E). The nonteacher completed the instructor questionnaire at the conclusion of the instruction and returned it by mail.

---

[13] While this method of dividing the class was not random in the strict sense, it was believed that it would preclude any biased selection of students on the part of the regular teacher. We thought more complicated randomization procedures might be confusing to the teachers.

RESULTS

Analysis. The analysis of the data was by classroom unit, i.e., the group of pupils taught by a teacher or nonteacher. Because of the procedure by which the groups had been formed, namely, two small classes split from a larger class, it was decided to treat them as correlated units. The contrasts, therefore, would be by teacher versus nonteacher for each original classroom unit. Essentially, then, 13 separate comparisons could be made with the mean performance of each group of pupils constituting the datum to be contributed to the analysis. Although parametric statistical procedures could have been employed in the analysis of such data, the sign test was considered acceptable in view of the prediction that the experienced teachers would greatly out-perform their nonteacher counterparts.

The principal analysis, of course, centered on pupil performance on the achievement tests. The tests were initially scored so that a top score of 50 was attainable. This involved a weighting decision on certain items where more than one answer could be correct. Each of these answers was given fractional credit, for example, if a three answer multiple-choice item was involved, each answer was counted one-third. This simple, gross scoring method was considered to be a rough check on the performance of the pupils on the achievement tests.

A more equitable, but more involved, scoring method was then undertaken so that an equal contribution would be made by the various test items for each of the 13 specific instructional objectives in the resource unit. Since the participants had been told to accomplish as many of the objectives as they could, it seemed appropriate to devise a scheme whereby the items related to each objective would make an equal contribution to the total test score. This was accomplished by determining the number of items related to a particular objective and then multiplying that number by a numerical quantity so that the resulting product would equal ten. A perfect score on the test, therefore, would have been equal to 130. In Table 3 the complete breakdown of items per objective, along with the multiplication factor used for each subset of items, is presented. The reader interested in seeing the particular items which were related to the objectives can do so by referring to the unit objectives and post-test found in the Appendices. For convenience, this scoring scheme was referred to as the "equalized objective" scoring method.

Table 3. Test Items Per Objective and Multiplication Factors Used in Deriving "Equalized Objective" Test Scores.

Objective Number	Test Items	Number of Items	Multiplication Factor
1	49-58	10	none
2	59-68	10	none
3	9,10	2	5.0
4	21-25	5	2.0
5	41-43	15	.67
6	11-20	10	none
7	1-8	8	1.25
8	26-30	5	2.0
9	31	1	10.0
10	32,33	2	5.0
11	34,35	2	5.0
12	44-48	5	2.0
13	36-40	5	2.0

Test Results. Contrary to predictions, the teachers did not perform better than the nonteachers when pupil achievement of prespecified objectives was the criterion. Using either the simple, gross scoring method of the "equalized objective" scoring method, no systematic differences in favor of the experienced teachers occurred. Sign tests yielded nonsignificant results. Using the simple scoring method (50 points perfect score), the mean and standard deviation for the teachers was 33.44 and 2.20; for the nonteachers the mean and standard deviation were 32.30 and 2.96. Using the "equalized objective" scoring method (130 points perfect score) the mean and standard deviation for the teachers were 88.64 and 6.55; for the nonteachers the mean and standard deviation were 88.17 and 7.92.

In Table 4 the means and standard deviations of the 26 subjects, using the simple scoring method are presented. As can be seen, in the 13 comparisons, eight favored the teachers and five favored the nonteachers.

A Kuder-Richardson (21) estimate of internal consistency was computed on the basis of the total n of 368 students (disregarding subgroups) for the simple scoring method. The Kuder-Richardson (21) coefficient yielded by this analysis was .75.

Table 4. Means and Standard Deviations of Post-Test (Simply Scored) by Teacher: Validation Study II

Class	Teacher		Nonteacher	
	$\bar{X}$	s	$\bar{X}$	s
1	31.49	4.36	36.89	5.57
2	33.14	6.46	33.79	5.95
3	33.07	3.78	31.26	6.60
4	35.07	5.17	35.85	7.00
5	37.78	6.87	34.47	5.60
6	34.61	6.70	30.75	10.89
7	28.41	6.57	34.54	5.42
8	34.51	6.40	30.82	7.70
9	35.86	4.14	34.76	5.70
10	32.43	4.57	27.63	4.05
11	29.25	6.97	30.16	6.67
12	31.79	4.83	31.13	6.77
13	31.79	6.09	27.86	7.27

In Table 5 the means and standard deviations of the 26 subjects, using the "equalized objective" scoring method are presented. In the 13 comparisons, seven favored the teachers and six favored the nonteachers.

Table 5. Means and Standard Deviations of Post-Test (Equalized Objective Scoring Method) by Teacher: Validation Study 2

Class	Teacher		Nonteacher	
	$\bar{X}$	s	$\bar{X}$	s
1	86.46	12.89	98.59	13.26
2	88.56	16.64	91.70	15.01
3	89.69	10.55	85.74	16.54
4	93.46	16.26	100.75	11.15
5	101.21	18.62	94.11	18.19
6	92.94	19.01	87.55	18.76
7	78.31	15.66	90.74	11.85
8	92.85	14.89	83.91	18.95
9	94.77	11.60	93.27	16.43
10	87.61	12.78	72.24	13.04
11	79.02	19.83	84.67	15.95
12	85.38	11.83	85.40	17.42
13	82.10	15.26	77.60	14.87

The performance of the subjects by objective was of considerable interest. The possibility existed, of course, that one group of subjects might be emphasizing certain objectives rather than others. Perhaps, for example, the nonteachers were focusing on some of the easier objectives rather than the more difficult, time-consuming objectives. While this might not show up in total score differences, it would be revealed in a comparison of the two groups' performance by objective. Accordingly, in Table 6 the results of 13 contrasts, i.e., teacher versus nonteacher, for all of the objectives are presented. The complete means and standard deviations of each teacher's and nonteacher's pupils are presented in 13 additional tables in Appendix G.

Table 6. Contrasts of Pupil Performance Favoring Teachers and Nonteachers by Objective

Objective Number	Contrasts Favoring Teacher	Contrast Favoring Nonteacher	Ties	P*
1	10	3	0	.10
2	8	5	0	NS
3	5	8	0	NS
4	7	6	0	NS
5	5	8	0	NS
6	7	6	0	NS
7	5	8	0	NS
8	5	8	0	NS
9	5	6	2	NS
10	6	5	2	NS
11	1	8	4	.05
12	5	8	0	NS
13	7	6	0	NS

*Sign test probability, two tailed

As can be seen in Table 6, six contrasts of objectives favored the teachers and seven favored the nonteachers. Most of these differences, however, failed to achieve statistical significance. Certainly there was no systematic difference evidenced by this type of analysis.

Affective Data. Responses to the pupil questionnaire and teacher questionnaire were converted to numerical scores by a simple coding system. For instance, if pupils were asked whether they were "extremely interested, interested, neutral, bored, or extremely bored" by the topics covered in the unit, values of 1, 2, 3, 4, and 5 were given to their responses. On the pupil questionnaire data means were computed for each subject's group of pupils.

No systematic differences between the teacher group and the nonteacher group were revealed on either the student questionnaire data or the teacher questionnaire data. Sign tests failed to detect any significant differences between the two groups on any of the measures.

DISCUSSION

Results of both validation studies failed to confirm the prediction that experienced teachers would promote significantly better achievement of given instructional objective than would nonteachers. How can this be explained? Does it mean that the preformance test approach to the assessment of teacher competence is unworkable?

Some of the possible reasons for the results obtained in the initial validation study were compensated for by adjustments in the second study. We cannot, therefore, easily explain away the unfulfilled prediction on the basis of such explanations as "The study should have been conducted in a school setting," or "The nonteachers were too highly motivated." Nor can we readily dismiss the lack of differences between teachers and nonteachers because of a faulty measuring device. The internal consistency estimates were acceptable and there was sufficient "ceiling" for high learner achievement.

Indeed, in the second validation study the teacher group had several clear advantages over their nonteacher counterparts. They were familiar with the school setting, e.g., classroom facilities, resource materials, etc. They knew their students, having worked with them for approximately three weeks prior to the time the study was conducted. Couple these rather specific advantages with those which might be attributed to teaching experience (such as ability to attain better classroom discipline, ease of speaking before high school students, sensitivity to the learning capabilities of this age group, etc.) and one might expect the teachers to do better on this type of task. The big question is "Why not?"

Although there are competing explanations, such as insufficient teaching time, the explanation that seens inescapably probable is the following: Experienced teachers are not experienced at bringing about intentional behavior changes in learners. When it comes to a task such as that presented by the performance test in which they must promote learner attainment of specific instructional objectives, most experienced teachers are no better prepared than a person who has never taught. One need only speculate on the typical intentions of most public school teachers to realize why this is so. They wish to cover the content of the course, to maintain classroom order, to expose the student to knowledge, and so on. Rarely does one find a teacher who establishes instructional objectives prior to teaching, objectives clearly stated in terms of learner behavior changes, and then sets out to achieve them. Only recently, in fact, do we find many teachers who are even familiar with the manner in which instructional objectives are stated in operational form.

Lest this sound like an unchecked assault on the integrity of the teaching profession, it should be quickly pointed out that there is little reason to expect that teachers should be skilled goal achievers. Certainly they have not been trained to be; teacher education institutions rarely foster this sort of competence. There is no premium placed on such instructional skill; neither the general public nor professional teachers' groups attach any special importance to the teacher's attainment of clearly stated instructional objectives. Whatever rewards exist for the teacher in his typical school environment are not dependent upon his skill in promoting measurable behavior changes in learners. Indeed, the entire educational establishment seems drawn to any method of rewarding instructors other than by their ability to alter the behavior of pupils.

To the extent that this is true, the attempt to validate the performance test of teaching proficiency by contrasting the performance of teachers and nonteachers was ill-conceived. It was wishful thinking to believe that experienced teachers would do better. But because this validation scheme was injudiciously selected does not mean that the performance test approach is unworkable, nor does it mean that such tests cannot be validated.

This validation effort can be classified as an effort to supply construct validity evidence. Another, more reasonable construct validation effort could be based on a contrast between (a) instructors who had manifested measurable skill in promoting learner attainment of prespecified objectives and (b) instructors who had not manifested such skill. The initial group of instructors could be trained on comparable, but unrelated, teaching tasks until they could show that when presented with instructional objectives specified in terms of learner behavior, they could accomplish such objectives. Then both the skilled and unskilled group could be given a performance test such as the social science instrument described in this report. The prediction would be, of course, that the skilled instructors would out-perform the unskilled instructors.

To persist in attempting to validate this approach to the assessment of teaching competence implies that one is in sympathy with the premise that the teacher should be an efficient behavior changer. It seems that this general conception is subscribed to, implicitly, by the majority of our citizens. Parents don't send their children to school with the expectation that nothing will happen to them. But the clarity with which these changes are described prior to instruction, and measured after instruction, is the area that needs improvement.

With increased clarity regarding the changes we wish to promote in learners, the general performance test strategy for the assessment of teacher competence still seems more than an acceptable method, it seems to be the acceptable method.

CONCLUSIONS

The principal conclusion of this project was that the participating teachers were not, in two separate instances, able to perform better than nonteachers with respect to their ability to promote learner attainment of prespecified instructional objectives. Obviously, generalizations beyond the types of teachers and nonteachers involved in the investigation, as well as the teaching task and pupils, should be undertaken cautiously. The explanation offered for these results was based on the teachers' lack of experience in achieving preset behavioral changes in learners.

An additional conclusion should also be mentioned. At the earliest conception regarding the use of these performance tests as measures of teaching competence, it was assumed that we could indeed develop a sufficiently sensitive index of a particular teacher's ability (to accomplish instructional objectives) so that this measure might be used as a rating of individual instructors. The idea, for example, of using results of one or more performance tests in merit rating schemes did appear possible. Granted that problems of variability within different teachers' classes existed, it was thought that this might be controlled, or at least adequately compensated for, through sensitive weighting procedures or simplified forms of statistical adjustment. It now seems that in light of the grossness of the measurement device, we shall be very pleased even if the performance tests are suitable for use only with groups. In other words, it will be a sufficient advance to develop a reliable group criterion measure which could be used in many educational situations, for example to assess the efficiency of certain teacher education training programs.

It is recommended that further validation efforts be directed toward the performance test approach to the assessment of teacher competence. As suggested earlier, one reasonable method of pursuing this task would be to contrast the performance of those instructors who, after training, manifested ability to accomplish instructional objectives with instructors who had not manifested such ability.

Several possible procedural modifications in the performance test instruments may be beneficial in future investigations. First, perhaps the quantity of "possible learning activities" provided the teacher should be reduced or eliminated altogether. It may be too structuring to present a number of such activities, even though some were designed to be less relevant than others. Perhaps only the content needs to be described, leaving the teacher completely on his own to devise instructional procedures. A second revision in the procedure might be made by giving the teacher some choice in the objectives he chooses, rather than prescribing only one set of such goals. Although this would require considerable additional work to develop comparable objectives (and criterion items) from which instructors could choose, there may be some attitudinal advantages to be derived from such a change.

Finally, to the extent that the performance test strategy focuses the attention of educators on the ends of instruction (i.e., post-instruction behavior changes in learners), rather than instructional means (i.e., teaching procedures) its ultimate impact should clearly be beneficial.

SUMMARY

A project was undertaken to develop and, hopefully, validate a heretofore untried method of assessing teacher competence, namely through the use of a performance test. The performance test was designed to function in the following way. Teachers were presented with a list of specific, operationally defined objectives for a particular topic and directed to teach to the objectives. Following the instructional period, students were tested on the behaviors stated in the objectives. Teacher competence was judged in relationship to the way their students performed on the criterion test. An attempt to validate this method of measuring teacher effectiveness involved contrasting the results produced by experienced teachers and nonteachers on a performance test dealing with research methods in the social sciences.

Two separate contrasts were conducted, the first involving six professionally trained, experienced student teachers versus six housewives for a six hour teaching period. The second involved 13 regularly credentialed teachers and 13 college students for a four hour teaching period. In neither contrast did the teachers perform significantly better than the nonteachers.

The results were interpreted as indicating that the experienced teachers were not more experienced than the nonteachers in promoting learner achievement of previously established instructional objectives. An alternative approach to validating the performance test strategy was discussed along with possible procedural modifications in the approach.

APPENDIX A

EXPERIMENTAL INSTRUCTIONAL UNIT

SOCIAL SCIENCE RESEARCH METHODS

An Experimental Instructional Unit
Prepared under the Direction
of
Eva L. Baker

Version I

January, 1967

Department of Education
University of California
Los Angeles

TABLE OF CONTENTS

Introduction A-1

Section One: Objectives
  General Objectives A-2
  Specific Objectives and Sample Test Items A-3
    General Introduction A-3
    Components of Research Studies A-4
    Hypothesis Formulation A-6
    Data Collection A-6
    Data Analysis A-6
    Interpretation of Results A-10

Section Two: Content Guidelines
  Preface A-12
  General Introduction A-13
  Social Scientists and Their Contributions A-13
    Possible Instructional Activities A-16
  Glossary of Social Science Terms A-17
    Possible Instructional Activities A-20
  Components of Research Studies A-21
    Hypothesis A-21
    Procedure A-21
    Selection of Population and Sample A-22
    Description of Sample or Objects A-22
    Data Collection A-22
    Data Analysis A-23
    Statement of Results A-23
    Interpretation of Results A-24
  Hypothesis Formulation A-24
    Possible Instructional Activities A-25
  Data Collection A-26
    Observations A-26
    Interviews A-27
    Questionnaires A-27
    Projective Tests A-28
    Personal Reports A-28
    Mass Communication Records A-29
    Experiments A-29
    Possible Instructional Activities A-30
  Data Analysis A-31
    Statistical Methods A-32
      Range
      Mode
      Median
      Mean
      Graph
    Authentication A-33
      External Authentication
      Internal Authentication
        Possible Instructional Activities A-34

TABLE OF CONTENTS (continued)

Interpretation of Results A-35
    Statements Supported or Contradicted by the Facts Given . . A-35
    Research Errors A-35
        Sampling Error A-35
        Confounding Variable A-36
        Authentication Error A-36
        Inadequate Data for Conclusion A-36
            Possible Instructional Activities A-37

Section Three: Resource Materials

Introduction A-38
Section One: Research Studies A-39
Section Two: Data Collection A-74
    Projective Tests A-76
        Thematic Apperception Tests A-77
        Confounding Variable A-79
    Experimental Design A-81
Section Three: Data Analysis A-83
    Statistical Summary Measures A-83
        Mode A-83
        Median A-84
        Mean A-85
    Uses of Central Tendency A-86
        Appropriate Uses of Averages A-86
        Range A-87
        External Authentication A-89
            Carbon-14 Dating A-90
        Internal Authentication A-91

SECTION ONE: OBJECTIVES

INTRODUCTION

The topic of the unit you are about to teach is "Social Science Research Methods," and it focuses its attention on the research techniques social scientists often use. The unit is organized in the following way:

In the first section you will find a list of general objectives to be achieved in the unit, followed by very specific objectives which you are to teach to your class. Try to get your students to achieve as many of these objectives as possible in the time you have available. You may find, however, that there are too many objectives for your particular class to learn. In that case, concentrate on the objectives you feel are most important; although you should attempt to teach as many as you can. Sample test items like those which will measure the objectives are provided after each specific objective.

The second section, entitled "Content Guidelines" is very important. It contains all the information you will need to help your students accomplish the objectives. Be sure to read this section, for certain concepts have been defined in particularly limited ways to make the unit suitable for this grade level. Included in the "Content Guidelines" is much of the actual material used to write the test questions for each of the objectives.

The "Content Guidelines" section is organized according to major components of research studies. In addition, it contains a glossary of terms (only the asterisked items of which will be tested) and a section of brief sketches about famous social scientists and their contributions to their disciplines. Following each explanation of a research procedure is a list of suggested classroom activities you might wish to consult. Many more activities than you could possibly use are included, and, of course, you may choose not to employ any of them. They are provided in order to make your instructional planning easier.

The last major section of the unit is labeled "Resource Materials," and in a very real sense is quite optional. Most of this section consists of summaries of actual research conducted in the social sciences. These studies were chosen to illustrate some of the research techniques treated in the unit. You may wish to draw on these for examples or give them to your students to study.

General Objectives

This unit seeks to develop in students certain abilities which will make it easier for them to understand how knowledge in the social sciences is systematically accumulated. Broadly, the unit will help students to

1. distinguish fact from opinion
2. analyze written material critically
3. interpret data
4. become familiar with some of the general procedures used in social science research
5. see the interrelationships among different social sciences
6. become acquainted with the proper documentation and reporting of social science research
7. learn what common pitfalls in social science research must be avoided
8. discover some of the famous social scientists of the past and learn about their contributions

In order to help accomplish these important goals, a list of specific objectives has been provided.

SPECIFIC OBJECTIVES AND SAMPLE TEST ITEMS

The following is a detailed list of objectives which you should try to achieve during the instructional period. These objectives are stated specifically in terms of your students' behavior. They are organized and labeled according to the area of social science inquiry with which they deal. Following each objective is a sample test item representing the kind of test questions which will be used for the objective. Try to achieve as many as you can.

GENERAL INTRODUCTION

Objective 1. Given a description of the key contribution of a social scientist, the student should be able to identify this social scientist from sets of four alternatives which include:

Thomas Carlyle
Auguste Comte
Sigmund Freud
Herodotus
Alexander von Humboldt
John Locke
Bronislaw Malinowski
Karl Marx
Margaret Mead
Plato
Ivan Pavlov
Adam Smith

Sample Test Item: DIRECTIONS: CHOOSE THE SOCIAL SCIENTIST WHO FITS THE FOLLOWING DESCRIPTION.

1. This historian produced the frontier theory of history. He suggested that the ever-expanding frontier with its limitless free land in America could be an explanation for the differences between Americans and Europeans.

[✓ a.] F. J. Turner
b. Thomas Carlyle
c. Bronislaw Malinowski
d. Alexander von Humboldt

Objective 2. Given alternatives, the student should be able to select the correct definition for asterisked social science terms given in the Glossary.

Sample Test Item: DIRECTIONS: CHOOSE THE ANSWER WHICH BEST COMPLETES THE FOLLOWING DEFINITION.

2. The facts or information collected and analyzed in a research study.

[✓ a.] data
b. variables
c. subjects

COMPONENTS OF RESEARCH STUDIES

Objective 3. From a set of alternative orders, the student should be able to select the proper ordering of the components of an experiment. These components are:

Hypothesis
Procedure
Selection of Population and Sample
Description of Sample
Data Collection
Data Analysis
Statement of Results
Interpretation of Results

Sample Test Item: DIRECTIONS: CHOOSE THE BEST ORDER OF STEPS FOR CONDUCTING AN EXPERIMENT.

3. Choose the best order of steps for conducting an experiment.

a. analyze data, state results, form hypothesis
b. state results, analyze data, form hypothesis
[✓ c.] form hypothesis, analyze data, state results
d. analyze data, form hypothesis, state results

Objective 4. Given a detailed description of a research study, the student should be able to identify from alternatives the various components of a research study which are listed above.

Sample Test Item: DIRECTIONS: READ THE FOLLOWING PARAGRAPH AND THEN CHOOSE THE BEST ANSWER FOR EACH OF THE QUESTIONS BASED ON IT.

4. Do teenage juvenile delinquents in the United States watch more crime programs on T.V. than law-abiding teen-agers? To test this question a sociologist gives questionnaires to 100 teen-agers who have been sentenced in the Los Angeles Juvenile Court for various offenses and 100 teen-agers who have never been in trouble with the law. (Each group of youngsters includes equal numbers of boys and girls between 13 and 19

years old.) The sociologist finds that the formerly delinquent teen-agers watch crime programs only slightly more than the other teen-agers. Having found such a small difference between the two groups, he hesitates to suggest any further hypothesis about the relationship of crime and T.V. crime shows. However, he does suggest that further study might show that juvenile delinquents spend more time watching all kinds of T.V. programs because they do not study or participate in sports as much as other youngsters.

The question that best serves as a hypothesis in this study is:

a. Do teen-age juvenile delinquents in the United States watch more T.V. programs than law-abiding teen-agers?
b. Do teen-age juvenile delinquents in America watch many T.V. programs because they have few other interests?
[✓ c.] Do teen-age juvenile delinquents in America watch more crime programs on T.V. than law-abiding youngsters of the same age group?

The data collection procedure used in this study is:

[✓ a.] questionnaire
b. interview
c. observation

The sample is:

a. 200 teen-agers in the United States
[✓ b.] 200 teen-agers in Los Angeles
c. 200 delinquent and law-abiding teen-agers in the United states.

Objective 5. Give a summary of a study, the student should be able to identify from a list of alternatives which, if any, of the above components is omitted.

Sample Test Item: DIRECTIONS: READ THE FOLLOWING SHORT PARAGRAPH AND CHOOSE THE SINGLE BEST ANSWER FOR THE QUESTION AFTER IT.

5. A historian who specialized in Mexican history discovered some old ammunition dumps in northern Mexico. He concluded that Pancho Villa, the bandit leader of a Mexican revolutionary movement, had great support throughout the countryside, but that the peasants did not participate in his movement very much because they were not familiar with modern weapons.

The necessary information left out is (none, or more than one answer may be chosen):

[✓ a.] hypothesis
b. selection of sample
c. data analysis (by authentication)
[✓ d.] interpretation of results
e. none of the above

HYPOTHESIS FORMULATION

Objective 6. Given a research hypothesis, the student should be able to classify it as either descriptive or causal.

Sample Test Item: DIRECTIONS: DECIDE WHETHER THE FOLLOWING QUESTION PRESENTS A DESCRIPTIVE OR CAUSAL HYPOTHESIS. MARK THE APPROPRIATE LETTER ON YOUR ANSWER SHEET.

6. What proportion of the people in New York were born in foreign countries?

[✓ a.] descriptive
b. causal

DATA COLLECTION

Objective 7. Given the name of one of the data collection procedures listed below, the student should be able to select from alternatives the major advantage or limitation of the technique.

Observations
Interviews
Questionnaires
Projective Tests
Statistical Records
Mass Communication Records
Experiments

Sample Test Item: DIRECTIONS: CHOOSE THE BEST ANSWER FOR THE FOLLOWING QUESTION. ON YOUR ANSWER SHEET FILL IN THE SPACE UNDER COLUMNS a, b, c, or d NEXT TO THE NUMBER OF THE QUESTION YOU ARE ANSWERING.

7. One limitation of using projective tests is:

a. the subjects easily find out the purpose of the test and do not behave naturally.
[✓ b.] the subjects may not make the same responses on repeated tests.
c. the tests cannot be given to young children or other subjects who cannot read or write.

DATA ANALYSIS

Objective 8. Given a description of a researcher's procedure, the student should be able to determine if internal authentication or external authentication is used to analyze the data.

Sample Test Item: DIRECTIONS: CHOOSE THE ONE BEST ANSWER FOR THE FOLLOWING SHORT DESCRIPTION.

8. A historian and an anthropologist were investigating the religious customs of the American Indians who lived just before the Civil War. After much searching they finally found a burial ground with grave markers which seemed to have religious significance. But the laboratory reported on the age of the markers showed that they had been buried for at least 500 years.

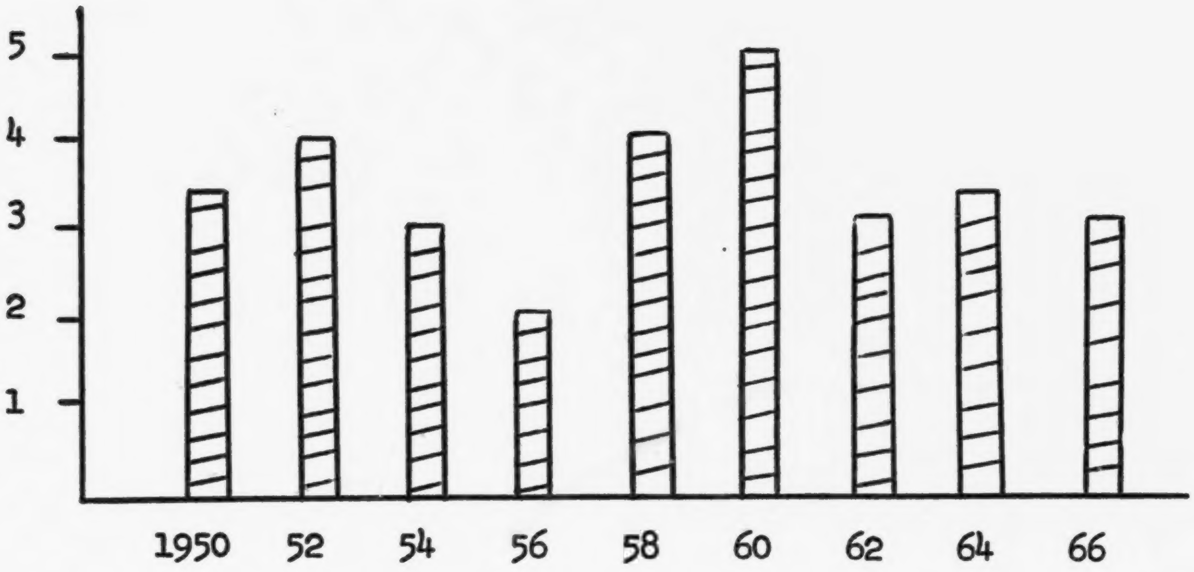
What authentication procedure did the social scientists use?

a. internal authentication
[✓ b.] external authentication
c. both
d. neither

Objective 9. Given a graph plotting simple data, the student should be able to identify from a list of alternatives the mode, mean, median, and range.

Sample Test Item: DIRECTIONS: LOOK AT THE FOLLOWING GRAPH AND ANSWER THE QUESTIONS ABOUT IT.

9.



Graph A

Students attending Baseball Games at Hillside Junior High School

What year is the mode in Graph A?

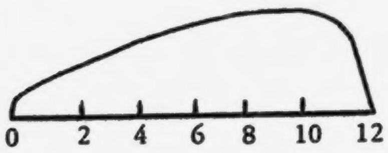
a. 1950
b. 1956
[✓ c.] 1960
d. 1966

Objective 10. Given either the mean and the range or the median and the range of a distribution, the student should be able to select from alternative sketches the distribution which most accurately depicts the statistics.

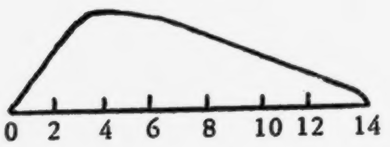
Sample Test Item: DIRECTIONS: LOOK AT THE FOLLOWING GRAPHS AND ANSWER THE QUESTION ABOUT THEM.

10.

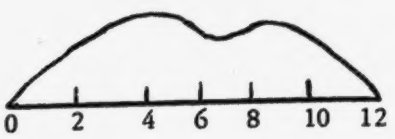
Graph B



Graph C



Graph D



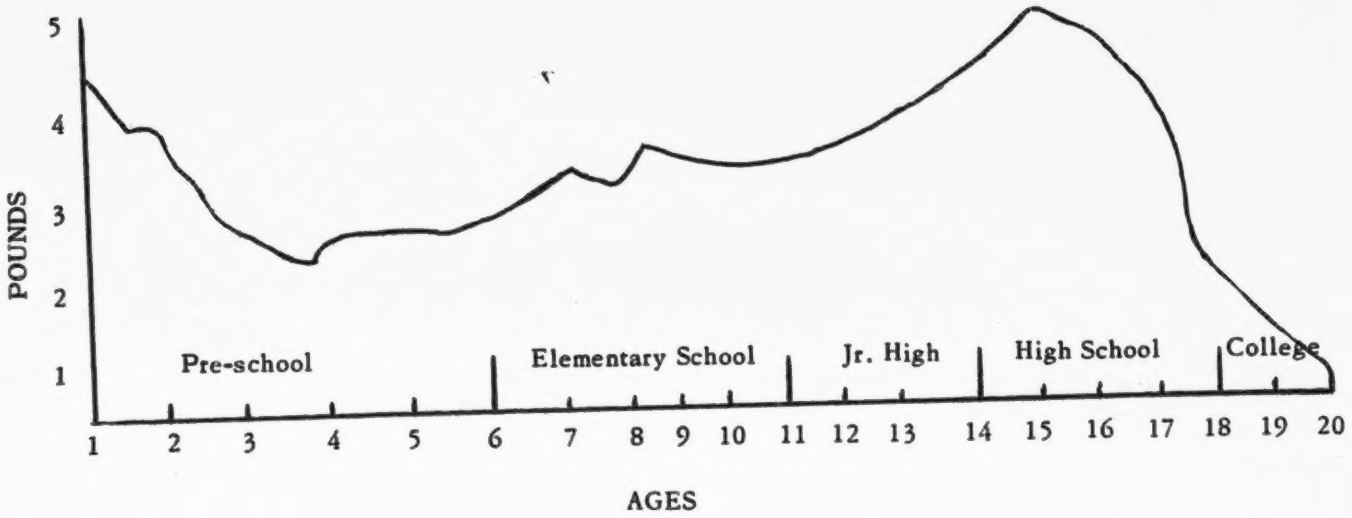
Which of the graphs above has a mean closest to 10 and a range of 12 units?

a. Graph B
[✓ b.] Graph C
c. Graph D

Objective 11. Given a bar, line, or circle graph, the student should be able to identify from alternatives the statements which most accurately describe the data in the graph.

Sample Test Item: DIRECTIONS: LOOK AT THE FOLLOWING GRAPH AND ANSWER THE QUESTION ABOUT IT.

11. GRAPH E
Weight Changes in Boys



The median of the distribution in Graph E is probably in the

a. pre-school years
[✓ b.] elementary school years
c. junior high and high school years
d. high school and college years

INTERPRETATION OF RESULTS

Objective 12. When presented with information gathered in a research study, the student should be able to decide if selected statements are supported, contradicted, or neither supported nor contradicted by the facts given in the study.

Sample Test Item: DIRECTIONS: READ THE PARAGRAPH BELOW. THEN DECIDE WHETHER THE FOLLOWING STATEMENTS ARE a. SUPPORTED, b. CONTRADICTED, OR c. NEITHER SUPPORTED NOR CONTRADICTED BY THE FACTS GIVEN. (USE ONLY THE INFORMATION GIVEN IN EACH PARAGRAPH. DO NOT GUESS ON THE BASIS OF YOUR OWN KNOWLEDGE OF THE SUBJECT.)

12. Easter Island has been a mystery to the outside world from the moment it was discovered. An anthropologist hypothesized that the Easter Islanders came from Peru and adapted Peruvian culture to their new island home. To test his assumption, he searched through the ancient ruins of Peru and Easter Island and consulted previous studies about the two areas. He was hoping to find a number of similarities between the two cultures. During his investigation, he came across small stone statues in both Peru and Easter Island that seemed to represent gods. The use of stones in buildings in Peru and on Easter Island was also somewhat alike. In historical studies he discovered some rather general similarities between the religious beliefs of the two peoples. Because he could find no definite similarities between Easter Island and Peruvian cultures, he concluded that the Easter Islanders did not come from Peru but from some other part of South America.

The Easter Islanders came from South America

a. supported by the facts given
b. contradicted by the fact given
[✓ c.] neither supported nor contradicted by the facts given

There were no similarities between Peruvian and Easter Island cultures

a. supported by the facts given
[✓ b.] contradicted by the facts given
c. neither supported nor contradicted by the facts given

Objective 13. When given a description of a research study, the student should be able to identify from the following list the error the researcher has made:

Sampling Error
Confounding Variable
Authentication Error
Inadequate Data for Conclusion

Sample Test Item: DIRECTIONS: AFTER HAVING READ THE PARAGRAPH ABOVE, CHOOSE THE SINGLE BEST ANSWER FOR THE FOLLOWING QUESTION.

13. The error that the anthropologist made was

a. sampling error
b. confounding error
[✓ c.] inadequate data for conclusion
d. none of the above

SECTION TWO: CONTENT GUIDELINES

These Content Guidelines are designed to provide you, the teacher, with all necessary information so you can teach your students to answer the test questions. The information is written for your use. If you wish to provide your students with this material you will probably have to simplify some of the language. For purposes of simplicity (so the unit could be understood by students at this grade level), certain concepts have been defined in unusually limited ways. Although you might not agree with our definitions and may recognize that many of the topics are much more involved than we have presented them, please accept the explanations in the unit and use them when teaching your students. Provided with each description of a social science research procedure are possible classroom activities or assignments which you might wish to give your students. Not all of these activities would normally be employed during the course of this unit, but they are included so that you may use them in your instructional plans if you desire. Do not feel obliged to use any of these activities when teaching the unit. We are anxious to have you use whatever procedures you wish, organized in the instructional sequence which seems best to you. The only important thing is for you to try to achieve the specific objectives described in Section One. To put it another way, the ends are given, the means are entirely up to you.

GENERAL INTRODUCTION (SPECIFIC OBJECTIVES ONE AND TWO)

As an introduction to the social studies unit, we are presenting brief descriptions of the key contributions of some of the more famous social scientists and a list of terms commonly used in the social science disciplines. The variety of contributions made by social scientists illustrates some of the differences and similarities among the various social studies fields. Additional information about noted social scientists can be found in the resource materials in Section Three.

SOCIAL SCIENTISTS AND THEIR CONTRIBUTIONS

Thomas Carlyle (1795-1881)

Thomas Carlyle used history as a tool to dramatize his philosophy rather than to make an objective record of past events. His great historical works, The French Revolution and The History of Frederick, Called the Great, reveal his beliefs that dictatorships are good if they are successful and that democracies cannot succeed unless the people are educated. His tirades against democracy and his extravagent praise of dictatorship are not an indictment of the common man as a man. He was convinced that with education the masses could gain the wisdom necessary for leadership. He urged the establishment of universal education and inspired the enactment of welfare legislation.

Auguste Comte (1798-1857)

Auguste Comte, who coined the word "sociology," is recognized as the founder of the discipline. In his attempt to use sociology as a tool for rational social reform, he stressed a scientific method (observation, experimentation, and comparison) of investigating social relationships. Although few of his principles have survived, he made an important contribution to the field by synthesizing the social thought of the century preceding him.

Sigmund Freud (1856-1939)

Sigmund Freud is most famous for developing the psychoanalytic method of psychology. He originated the method of free association which he used to study 1) the effect of the unconscious mind on the conscious mind and action, 2) the importance of defense mechanisms which modify or repress instincts in abnormal and normal minds, 3) personality structure, 4) the motivating force of instincts in humans, 5) the importance of sexuality in children, and 6) the meaning of dreams.

Herodotus (c. 485-425 B.C.)

Herodotus is known as the father of history. His great prose epic, *Histories*, is a masterpiece of literature and careful historical research into the background and the battles of the Greco-Persian War of 490-497 B.C. In his words, he intended to "prevent the great and wondrous deeds of Greeks and Barbarians (non-Greeks) from losing their due meed of glory and withal to record what were their grounds of feud." From hearsay, offical records, and his own conjecture he gathered details of the history, the customs, and the great men of the period.

Alexander von Humboldt (1769-1859)

Alexander von Humboldt was a pioneer of scientific geography who travelled to different parts of the world, particularly Central and South America, to observe geographic phenomena. He was not content to collect and describe vast amounts of unrelated data as the earlier geographers had done. Instead, Humboldt used his knowledge of physiological and biological processes to develop systematic classifications, comparative descriptions, and scientific measurements of geographic conditions. He was one of the first to correlate the distribution of plant life with altitude and temperature. As he journeyed through Central and South America, he noted the languages and cultures of the area, determined the density of the population, and studied the influence of geography on social and political life.

John Locke (1632-1704)

The political liberalism and constitutional government espoused by John Locke in *Two Treatises of Government* have had great influence on modern political ideas and institutions. He did not agree with many of his predecessors that man is basically evil, but believed that man in the state of nature could be good or evil. Also, all men had certain basic rights in the state of nature that should not be denied by political rulers. Locke's concept of these inalienable rights (life, liberty and property) was incorporated into the American *Declaration of Independence* and *Constitution* (life, liberty and pursuit of happiness). He promoted the idea that the consent of the governed in a social contract is necessary to establish and to maintain a legitimate government.

Bronislaw Malinowski (1884-1942)

Bronislaw Malinowski, a British anthropologist, revolutionized field research in his four year in-depth study of Melanesia. He was a pioneer of modern social anthropology and was one of the first anthropologists to go to the exotic, primitive areas of the world to study men as members of whole societies. He felt that he could understand man in society best by analyzing small, simple social groups where man's needs are met by the simplest, most basic social institutions. Through his books, such as

The Family Among Australian Aborigines and A Scientific Theory of Culture, he developed his "functional" theory of anthropology, which, in summary, is the social institutions of primitive people are not outlandish or irrational developments but function, in different ways, to serve human physical and social needs.

Karl Marx (1818-1883)

Karl Marx, in Capital and the Communist Manifesto, expressed the theory that economics determines the course of history. He divided history into periods according to their dominant economic systems and argued that history progresses inevitably from one period to the next. Each period would end as its internal contradictions cause it to fall and make way for the next stage. The ultimate stage in historical development, according to Marx, would be that of communism. Communism would support a classless society which would no longer be torn by class struggle. Marx's greatest error was assuming that observation of a few instances, (for example, the evils which existed in the early stages of capitalism), could support scientific natural laws of economics. He was unable to see far enough beyond his own time to foresee the improved conditions of workers in the advanced capitalist countries.

Margaret Mead (1901-)

Margaret Mead, a famous American anthropologist, has been well known for her psychological approach to the study of human societies. Her books, Coming of Age in Samoa and Growing up in New Guinea, revealed the great extent to which personality is influenced by culture. As these studies indicate, Miss Mead spent many years engaged in field work in Samoa, New Guinea and other primitive areas. But she did not confine herself to studying small, simple societies. In her later studies of national character, such as Keep Your Powder Dry and Soviet Attitudes Toward Authority, she attempted to analyze the interrelationship of culture and personality in large, complex nations, such as the United States and the Soviet Union.

Plato (427-347 B.C.)

Plato's Republic is considered to be the first work of political science. In this political study, Plato wrestled with philosophical problems such as justice and the good life. He based his political theory on a study of human nature and was thus able to recognize in ancient Greece the basic challenges that have always faced men in society. He introduced the revolutionary notion that politics should be analyzed rationally rather than accepted through faith and fear. By rational analysis, the "right" kind of government, one which would promote the good life and justice, could be created.

Ivan Pavlov (1894-1936)

One of the most influential experimental psychologists was Ivan Pavlov, famous for his studies of conditioned learning in dogs and other animals. To condition a dog by classical conditioning, the animal is repeatedly given food immediately after a bell is rung. Soon the dog will expect food whenever he hears the sound of the bell and will begin to salivate when the bell is rung even if no food is presented. Pavlov's theory that even complex human mental processes can be explained in terms of classical conditioning has been widely criticized. But in stripping away some of the mystery surrounding mental processes, he helped generate acceptance of the idea that the mind and body are not separate but form a whole which can be studied by scientific methods.

Adam Smith (1723-1790)

Adam Smith was the prime exponent of the classical theory of economics, a justification for industrial capitalism which is still widely believed. In his well known book, Wealth of Nations, he claimed that national progress and wealth are best secured by freedom of private enterprise. He opposed government interference in the economy because of faith in a natural law which guides the free market process to keep supply, demand, and price in order. Smith believed that selfish interests are guided by an invisible hand to promote the general welfare.

POSSIBLE INSTRUCTIONAL ACTIVITIES

1. Students could be assigned an oral or written report on a famous social scientist to be given in class or handed in as a written assignment. If the school has adequate facilities, a class period could be spent in the school library collecting facts for these reports.

2. Concurrent with the unit, students might be assigned book reports relevant to the material in the unit, for example, Kon-Tiki, The Warren Report, The McCone Commission Report. These would be handed in at the end of the unit.

3. Students might role-play and have a debate between two or more famous social scientists, for example, Karl Marx and Adam Smith. This debate could lead to a general class discussion, with the class voting on which student presented his case most convincingly.

4. As part of a notebook for collection of material throughout the unit, the teacher could distribute dittoed material giving basic facts about the most famous social scientists. After noting basic contributions of each man in either a lecture form or by a general discussion, the teacher could require memorization of these facts and quiz the students briefly the following day.

GLOSSARY OF SOCIAL SCIENCE TERMS

analysis - an examination of a complex whole in which the nature of its parts and their relations become clear.

*anthropology - the study of man's physical nature (biological evolution and racial divisions) and his cultural nature (social organizations, languages, tools, and works of art and architecture).

authentication - the determination of whether information is accurate or an object is genuine. External authentication tells whether the information or the object has the physical properties (age, etc.) which it is supposed to have. Internal authentication reveals whether the content of verbal or written reports is reasonable. Authentication errors occur when no attempt, or an inadequate attempt, is made to verify data.

*carbon-14 dating - an approximate means of dating organic material. The radiocarbon atoms in living material start decaying at a uniform rate as soon as the material dies. Thus, when a portion of the material is burned and reduced to pure carbon, the amount of radiocarbon remaining in the substance indicates how old it is.

cause and effect relationship - a relationship in which a particular cause should produce a particular effect or result. The relationship can be proven only in an experimental study.

contradict - to deny the truth of a statement, especially by establishing that the truth is just the opposite of that suggested.

culture - the way of life of a particular group of people. It includes the whole range of human activities, not just the artistic or literary activities to which it usually refers.

*datum, data - the facts or information collected and analyzed in a research study.

distribution - the position, arrangement, or frequency of events, people, measures, etc., in an area or time period. In statistics, distribution often refers to a set of measurements arranged in an order from lowest to highest.

*economics - the study of the production, distribution, and sharing of resources. Economists are especially interested in the parts of society involved in agriculture, industry, and trade.

*experiment - a carefully planned and controlled research study in which a certain condition is varied in order to determine whether or when the condition causes certain effects. Experimental studies deal with causal hypotheses while nonexperimental studies deal with descriptive hypotheses.

---

*See Specific Objective 2.

frequency - the count of the number of times a measurement, an event, or a condition, etc., occurs. A frequency distribution is a set of measurements arranged from lowest to highest (or vice versa) which is accompanied by an indication of the number of times each measurement in the set occurred.

*generalize - to assume that the characteristics of a small group or a few specific facts probably apply to (are characteristics of) the entire group or a larger body of related facts.

*geography - the study of places on the surface of the earth. The differences and similarities of natural and cultural features are examined in order to determine the causes and effects of geographic conditions.

graph - a picture or a chart that shows the relationship between variables.

*history - the interpretive or descriptive record of events according to time.

*hypothesis - a tentative assumption that guides the collection, observation, and interpretation of facts. A descriptive hypothesis describes characteristics while a causal hypothesis assumes cause and effect relationship.

inadequate data for conclusion - a conclusion that goes beyond the information revealed in the study or that describes a cause and effect relationship in a nonexperimental study.

interpretation of results - the explanation, conclusion, prediction, or generalization based on the results of a study.

interview - a method of collecting data in which a subject is asked oral questions by a researcher who is physically present.

mass communication records - written materials (such as newspapers, magazines), radio and T.V. programs, and movies, which are intended to inform, persuade, or entertain large numbers of people. These records are not originally produced for research studies but may be collected as research data.

*mean - the average of a series of measures which is found by adding the measures and then dividing the sum by the total number of individual measures.

*median - the middle point in an ordered series of measures above and below which are an equal number of measures.

*mode - the most popular or frequently appearing measure in a distribution of measures.

*observation - a method of collecting data in which relevant information about anything that can be seen, especially human subjects and events, is noticed and recorded by a trained observer. Structured observation is guided by clearly and specifically defined categories. Unstructured observation is bound by few guidelines and is not limited to narrow, specific categories of behavior.

personal reports - written materials such as autobiographies, letters or diaries in which the authors describe their personal experiences. These documents are not produced for a research study but can be used as sources of research data.

*political science - the study of the processes, forms of organization, institutions, purposes, and interactions of governments.

*population - the total number of possible people, animals, objects, events, measurements, etc., which the researcher plans to study, usually by means of a small sample of the whole.

*projective test - a method of collecting data about attitudes, emotions and other aspects of the personality. Subjects are free to interpret, react to, or tell stories about pictures, symmetrical forms, a set of dolls and other things which could represent different things or situations to different people.

*psychology - the study of human or animal behavior. It focuses particularly on the mental adjustment of individuals to their environment and mental processes such as thought and emotion.

*questionnaire - written sets of questions which are given to subjects whose written answers are the data for research studies.

*range - the difference between the highest and lowest measures of a group of measures.

relationship - the connection between or some other quality about two or more conditions that are considered together. For example, an associational relationship is one in which two conditions are associated with each other or occur together. A causal relationship is one in which one condition causes another.

research error - a flaw such as improper procedure or overlooked influences which distorts or invalidates the results of the research study. The error can usually be avoided by a careful researcher.

research study - a careful investigation or experimentation aimed at discovering and interpreting facts to test a hypothesis.

*sample - a small part of a population (larger group) which is directly investigated in a research study. It should be representative enough and large enough to accurately reflect the population. A representative sample can be found by random sampling (chance selections of subjects). Sampling error occurs when a sample is randomly selected.

social science - deals with the institutions and functions of human society and with the interrelationships of the individual members of society. It includes the studies of anthropology, economics, geography, history, political science, psychology, and sociology.

*sociology - the study of social relationships, the consequences of those relationships, and the process of social change. It focuses on the beliefs, values, and language of the group rather than the individual.

statistical methods - mathematical procedures for collecting, analyzing, interpreting, and presenting groups of measurements. They can show causal, associational and proportional relationships between variables. The mean, median, and mode indicate the relationships of measures to the center of a distribution. Graphs present relationships among variables in a pictorial way.

subject - the person or animal which is being studied in a research project.

systematic - methodical or orderly procedure.

*variable - a condition (such as temperature) which can change certain aspects of its character without losing its basic identify (temperature). An independent variable is a condition which a researcher deliberately varies or which he finds already varied in nature. A confounding variable operates in addition to the variable that is being tested in an experiment and distorts or hides the relationship of the independent variable to a dependent or criterion variable.

POSSIBLE INSTRUCTIONAL ACTIVITIES

1. A panel discussion or debate could counter on such questions as "Can Social Science really be 'scientific'?"

2. A dittoed sheet with the glossary terms and definitions could be utilized by the teacher for a general class discussion. Perhaps several terms could be memorized each day as a beginning class activity, with the teacher giving practical illustrations on how these terms are used in the social sciences. The dittoed material should be kept in the student's notebook for future references.

3. A spelling test could be given on the glossary terms after the students had entered all the terms in their notebooks. Results of this test and of all quizzes might well be recorded by the students and used later in the unit to prepare graphs (which are treated in another section of the unit).

4. Students could use the glossary terms in proper sentences, either to hand in for correction or to keep in their notebooks.

COMPONENTS OF RESEARCH STUDIES

(SPECIFIC OBJECTIVES THREE THROUGH FIVE)

On the following pages the components of research studies are described in the order they should appear in a study. All of the components except the "Selection of Population and Sample" can be found in the studies presented later in Section Three's Resource Materials. "Selection of Population and Sample" will be included only in those studies in which generalizations are made from a sample of selected cases to all possible cases of a particular type. In other words, this component would be included in a research study which deals with one or a few people, objects, or events meant to be representative of a larger group of people, objects, or events.

The following sequence of descriptions should be used for objectives 3, 4, and 5, which require students to 1) select the proper ordering of the components of an experiment, 2) identify the components of a research study, and 3) identify the information omitted from a research study. Some of these topics are treated in greater detail later in this section. For instance, specific objective six (6) deals with hypothesis formulation, hence, the formulation of hypotheses is treated later in addition to the following paragraph.

HYPOTHESIS

The statement of the hypothesis tells what the researcher is trying to find out; it is the tentative assumption which will guide the researcher's collection and interpretation of data. In an experimental study, a causal hypothesis is made. It asks if a particular characteristic or event is the cause of another condition. In a nonexperimental study, a descriptive hypothesis is used. The descriptive hypothesis asks if an object, a person, or a situation has a certain characteristic. It also asks if two or more conditions (such as red hair and freckles) are associated with (not caused by) each other.

PROCEDURE

The procedure is a plan describing how the researcher will test his hypothesis. An anthropologist might conduct a nonexperimental "field study" and travel to Thailand to find out how much rice the average Thai family eats. A historian might search through the Congressional Record to support his hypothesis that a certain senator is strongly in favor of civil rights legislation. To determine if the experimental treatment causes changes in subjects, a psychologist might use a "before-after design" in which one group of subjects is tested before and after the experimental treatment. A sociologist might use an "after-only design" in which no pretest is given and two groups of subjects are used.

SELECTION OF POPULATION AND SAMPLE

The population is the total number of possible people, animals, objects, events, measurements, etc., which the researcher plans to study. The sample is a small part of the population which will be directly investigated in the research study. A population and a sample are selected for a research study only if the researcher wants to generalize from the people (etc.) he is studying to other people (etc.) which he cannot study directly. For example, if a psychologist is studying the dating habits of American teen-agers, his population is all American teen-agers, and his sample is a selected group of American teen-agers which should represent the whole group.

The sample must be representative enough to reflect accurately the population of possible people, objects, events, or measurements.

A representative sample must include proper proportions of the various elements in a population. Representative samples may be found by random selection. Random selection involves choosing the number of people needed for the sample by chance. Each member of the population has an equal chance of being selected. For instance, a researcher may have a long list of people or measurements from which he chooses every tenth item. Also, the researcher could flip a coin to see which subjects he will include in his study.

DESCRIPTION OF SAMPLE OR OBJECTS

If the researcher is investigating the characteristics of a sample from a population, the sample must be carefully described. Suppose that the sample is a group of human subjects. Are they young or old, male or female, rich or poor? If the researcher is not trying to make generalizations from a sample, the kinds of subjects, objects, or events that he is studying must be clear. The data gathered from the verbal reports of people or documents produced by people cannot be interpreted properly unless the people as well as the documents and reports are described. In other words, all of the relevant characteristics of a sample or other objects or studies must be reliably noted.

DATA COLLECTION

The researcher chooses a data collection technique according to the nature of his hypothesis.

The hypothesis may impose a few limitations on the kinds of data a researcher can collect if the research area is new or broad. Suppose the hypothesis is that some ruins found in the Arabian Desert were built by invaders from tropical Africa rather than from the Mediterranean coast. The researcher would have to collect a very wide variety of data from the ruins, from the Mediterranean coast of Africa and from tropical Africa. The data might be written records, artifacts, skeletal remains, etc. On the other hand, if the hypothesis indicated a very narrow research area, such as the relation between the heights and weights in a fifth grade class, the researcher will have only a few measures to collect.

The hypothesis also indicates the subjects of the study which are another guide to the data collection. If a psychologist wants to question a large number of people, a questionnaire would be most economical. If he has few subjects and wants detailed information about their feelings and attitudes, projective techniques would be best. Historical studies based on past events must rely heavily on statistical records, mass communication records and other written accounts.

DATA ANALYSIS

In order to discover whether the data support his hypothesis, a researcher must use a form of data analysis.

Some data require statistical analysis. If a sociologist has hypothesized a cause and effect relationship, he may have to use statistical analysis to establish that the cause will usually produce a particular effect. An event must produce an effect with high probability before a cause and effect relationship can be inferred. If an economist is making comparisons of measures of income in France and Britain, graphs or charts of means, medians, and modes would reveal how the two incomes compare.

Other data require external or internal authentication. If an anthropologist wants to find out if a cave painting was made in a certain historical period, he will send a portion of it to be externally authenticated by Carbon-14 dating. A Civil War historian might authenticate a written description of the Confederate soldiers by determining whether the writer might have been strongly biased on the Union or the Confederate side of the war.

STATEMENTS OF RESULTS

After the data are analyzed, the researcher is able to state the results of his study. The statement of results will tell whether the hypothesis is supported. For example, "Is the assumed cause and effect relationship found?" "Are most New Yorkers foreign born?" "Is the arrowhead the oldest known artifact in North America?"

INTERPRETATION OF RESULTS

In interpreting the results of the research study, the researcher comes to a conclusion about the hypothesis. If he finds research errors, further research on the hypothesis is necessary. He may be able to make predictions from the results of his limited study and be able to generalize from his study to a larger area of investigation.

HYPOTHESIS FORMULATION

(SPECIFIC OBJECTIVE SIX)

In order to start his research, an investigator often formulates an hypothesis (a tentative assumption) about the material which he is going to study. Even though he cannot yet prove his hypothesis, the researcher must already know something about the subject he is planning to study or he would not be able to formulate the hypothesis. The hypothesis guides him both in the collection and observation of facts and in the interpretation of those facts. The hypothesis can be made either in the form of a statement which will be proved or disproved or in the form of a question to which (it is hoped) an answer will be found.

Social science research can be divided into two main categories: 1) nonexperimental studies concerned primarily with the description of people, things, and events; and 2) experimental studies concerned with cause and effect relationships. Thus, two different kinds of hypotheses are used to guide the two types of research: 1) descriptive hypotheses and 2) causal hypotheses.

Descriptive hypotheses are used in nonexperimental studies. One type of descriptive hypothesis asks whether an object, a person or a situation has a certain characteristic. This type of hypothesis is often employed by historians or anthropologists, as well as other social scientists. Descriptive hypotheses may also deal with the frequency of occurrences or associations among variables. They ask whether "something" occurs a certain proportion of the time, whether something tends to be accompanied by something else, or whether something is usually greater or less than something else. Frequency types of descriptive hypotheses are often used by sociologists or political scientists.

> Example: A historian thinks that many attitudes of Moses could be explained if it could be shown that Moses was an Egyptian rather than a Jew. He starts his research with the question: "Was Moses an Egyptian?" He then collects his evidence in such a way as to provide an answer to this question.

Example: A sociologist thinks there might be a relationship between watching crime programs on television and committing crimes, so he asks the preliminary question: "Do people who commit crimes watch more crime programs on television than people who don't commit crimes?"

Example: A political scientist wants to find out public opinion on the question of admitting Red China to the United Nations. He starts his research with the question: "What proportion of the population of the United States strongly opposes the admission of Red China to the United Nations?"

Causal hypotheses, on the other hand, ask whether a particular characteristic or event is one of the causes of another characteristic or event. Experimental studies are necessary to establish the existence of such cause and effect relationships.

Example: A psychologist has observed that disturbed children act very restless when placed in a room whose walls are red, but that they are calmer when they play in a room whose walls are blue. He starts his research with the question: "Does the color of the walls of a room cause changes in mood among disturbed children?"

Example: A schol psychologist wants to know what length of time to make the classes in high school. He starts his study with the question: "Does the length of the class period result in differences in student learning?"

POSSIBLE INSTRUCTIONAL ACTIVITIES

1. A class discussion could center around the use of hypothesis and assumption in everyday life of the students, pointing out the differences between scientific experiments and assumptions.

2. The teacher defines causal and descriptive hypotheses, reads a study from the Resource Materials, and discusses with the class whether the hypothesis is causal or descriptive.

3. After giving several examples, give the class a short quiz asking them to identify from a list read by the teacher those hypotheses which are causal and those which are descriptive. The teacher should discuss the answers to the quiz and give more examples.

4. The teacher could give the class a problem for research and ask them to make up hypotheses, one of which would serve as the basis for a causal study, and one for a descriptive study.

5. Students as a homework assignment could be asked to locate from the newspaper or other source a study which uses a causal hypothesis and one which uses a descriptive one.

DATA COLLECTION

(SPECIFIC OBJECTIVE SEVEN)

Among the ways data can be collected are: arranging for observations, conducting interviews, giving questionnaires or projective tests, investigating personal reports, statistical records, or mass communication records, and setting up experiments. A researcher chooses a method or a combination of methods of collecting data by considering the advantages and limitations of each collection procedure in terms of his hypothesis. The hypothesis indicates, for instance, whether the subjects can be approached directly with observation, interview, questionnaire, projective, or experimental techniques. If the subjects are not living, or if events are the focus of the study, personal reports, statistical records and mass communication records can be used.

OBSERVATIONS

Definition

Observation is a method of collecting data in which relevant information is noticed and recorded by a trained observer. The data may concern people, objects, events, or anything else which can be seen. The information should, of course, be relevant to the research purpose. Though the observation must serve a formulated research purpose (a specific end), the observation procedures (means) may be structured or unstructured depending on the degree of planning.

Structured observation is guided by clearly and specifically defined categories, while unstructured observation is not limited to preconceived categories. Observers have few guidelines to limit their collection of data; thus, the mass of unstructured data they collect might be used to suggest new research hypotheses.

Advantages

1. Behavior can be recorded as it occurs, not merely as it is recalled.
2. Subjects, such as infants, who cannot speak or write can be used.
3. In structured observation, observers can be trained to obtain the same kinds of information.
4. In unstructured observation, new research hypotheses may be suggested.

Limitations

1. The complexity and newness of unfolding events can impede accurate recording.
2. The presence of observers may change the behavior of the subjects.

INTERVIEWS

Definition

An interview is a method of collecting information in which a subject is asked oral questions by a researcher who is physically present. The data collected should be relevant to a formulated research purpose, but the questions may be either prepared in advance or improvised by the interviewer.

Advantages

1. All types of people can serve as subjects.
2. The interview can be very flexible. If necessary, the interviewer can interpret, repeat, or rephrase questions for the subjects.

Limitations

1. The presence of the interviewer may influence the answers of the subjects.
2. Interview data cannot be compared easily because interviews are so flexible and dependent on the reactions of subjects and the personalities of the interviewers.

QUESTIONNAIRES

Definition

A questionnaire is a written set of questions given to research subjects whose written answers are the data for an investigation. The questionnaire can be administered in person to individuals or groups. Questionnaires can also be mailed to individuals.

Advantages

1. Subjects feel they are under less pressure to answer in certain ways.
2. Those who give the questionnaire do not have to be specially trained.

3. The data collected can be compared easily by statistical methods because the instructions and order of questions are standardized.

Limitations

1. A large proportion of people do not answer and return questionnaires because of lack of interest, time, etc.
2. Many people whose reading ability is below average have difficulty in responding to questionnaires.

PROJECTIVE TESTS

Definition

With a projective test, subjects are free to interpret or react to deliberately ambiguous stimuli in their own ways. Examples are the ink blot Rorschach test, a picture about which many stories can be told (Thematic Appreception Test--TAT), a set of dolls which can be made to act in various ways, psychodramas, word association tests, etc.

Advantages

1. Subjects can answer freely and are not limited to a certain set of alternatives.
2. Subjects cannot change their responses to conform to what is expected of them because the ambiguous stimuli do not reveal the purpose of the test.

Limitations

1. Subjects may not make the same responses on repeated tests.
2. Comparisons among responses are difficult.

PERSONAL REPORTS

Definition

A personal report is a written document such as an autobiography, a letter, or a diary which focuses on the writer's personal experiences and is produced on his own initiative.

Advantages

1. The reports are not solicited by the researcher nor influenced by his biases.
2. The writer's words might supplement and corroborate other research techniques.

Limitations

1. It is hard for the researcher to recognize deliberately or inadvertently distorted or unclear information given in the personal report.
2. The reports cannot be compared easily because they are not uniform in content, mood, style, etc.

MASS COMMUNICATION RECORDS

Definition

Mass communication records are documents produced by a literate society intended to inform, persuade, or entertain people. Examples are newspapers, magazines, radio and television programs, and movies.

Advantages

1. They are not influenced by the researcher's bias.
2. Information about the differences and similarities of past and present societies can be obtained.

Limitations

1. Not all of the mass communication records that are desired from a particular time are available.
2. Mass communication records cannot be compared easily because viewpoints expressed may vary with the policy of the particular publication or medium, e.g., Time, New Republic.

EXPERIMENTS

Definition

An experiment is a research study that follows a carefully planned procedure to test a causal hypothesis. In an experiment the researcher varies certain conditions in order to determine variations in these conditions that cause certain effects. For instance, a researcher may vary some condition in the environment of a child, such as some aspect of the teacher's behavior. He then seeks to determine how this affects achievement as measured by a test. Both the causes and the effects are called variables. These variables can be environmental, physical, or psychological.

In a good experiment, the experimenter controls all conditions (confounding variables) that are likely to distort a result. An example of a confounding variable which could disrupt an experiment is a subject's awareness of the condition being tested. Thus, if speed-allowed-for-response were the variable being manipulated, the experiment could be ruined if the subjects were aware of the key variable and tried to respond as fast (or as slowly) as they could.

Advantages

1. It can be set up carefully so that potentially confounding variables do not exercise a confusing influence on the hypotheses under investigation.
2. A cause and effect relationship can be shown only through an experiment.
3. It permits replication whereby other, identical experiments can be conducted in order to verify the results of previous experiments.

Limitations

1. The situation is often artificial and may not be found in real life.
2. Subjects are often aware that they are being experimented on and thus may give unusual or "guinea pig" responses.
3. To maintain adequate control over irrelevant, confounding variables the time period for an experiment is often too short to obtain realistic results.

POSSIBLE INSTRUCTIONAL ACTIVITIES

1. Using one of the studies in the Resource File (for example, "Age as a Variable," or "Teen-agers and Their Dislikes," ask the students to identify orally the techniques of data collection used.

2. Discuss with the students the advantages and limitations of each collection technique. Then supply them with a research problem and discuss which technique would be most suitable and why.

3. Visit a local business for the purpose of observing their data collection procedures. If possible, have some individual who works in

this field visit the class and lecture to them about his work in data collection.

4. The students prepare their own study based on a topic supplied by the teacher. This study should use one or more of the collection techniques discussed. The data could be collected outside of class, or the students could role-play as investigators and collect data from their classmates during class time.

5. After a discussion on advantages and limitations of the various techniques, give the students a duplicated sheet on which he is asked to match the procedure with the advantages and/or limitations.

6. One day's discussion could center around the Nielsen ratings which are the basis for success or failure of television programs. After a report by the teacher or a group of students, other students could discuss the limitations and/or advantages of this data collection system.

7. Supply a topic for research or draw one from class discussion and have each student write the hypothesis, collect the data, etc. Results are presented in class and could lead to activities involving data analysis.

8. Ask someone from the Research Bureau of the Local School District to lecture to the class about the methods used in his department.

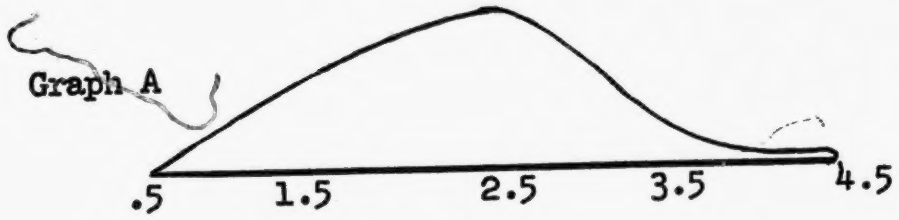
DATA ANALYSIS

(SPECIFIC OBJECTIVES EIGHT THROUGH ELEVEN)

Data analysis reveals whether data support the research hypothesis. If the hypothesis asks if a particular condition causes a certain effect, if a variable is associated with another variable, or if something occurs a certain proportion of the time, then statistical methods can be used to analyze the data. Graphs can often present quite clearly the relationships between variables which are uncovered in statistical analysis. If, however, the hypothesis asks whether a person, thing or event has certain characteristics, authentication procedures should be used. Authentication analysis shows whether a given datum, such as a historical document, is really what it appears to be.

STATISTICAL METHODS

Range



The range is the difference (in graphs, the distance) between the highest and lowest measures (test scores, height, etc.) of a group of measures. In Graph A, the range is between .5 and 11.5. That is, the range is 4 scores (4.5 - .5 = 4).

Mode

The mode is the most frequent or popular measure in a distribution of measures. On graphs such as the one above, the mode is found by dropping a straignt line from the highest point of the distribution to the measure line at the bottom of the graph. In Graph A the mode is 2.5.

Median

The median is the middle point in an ordered series of measures. The median cannot be found until the measures are ordered from the lowest to the highest (on a graph or list). An equal number of measures lie above and below the median. For instance, in the following set of scores, the median is 6: 4,4,5,6,7,7,7.

Mean

The mean is the average of a series of measures. It is found by adding the measures together and then dividing that sum by the total number of individual measures. For example, the mean of the following test scores would be 6.5: 9,8,8,7,7,6,5,2.

Graph

A graph (sometimes called a chart or a diagram) is a pictorial representation of the relationship between variables. Graphs can picture relationships in many different ways, depending on the nature of the data and the purpose of the graph. Some of the more popular forms of graphs are line graphs, bar graphs and circle graphs.

AUTHENTICATION

Social scientists often use research material from the past or present which has not been produced specifically for a research project. Examples of such materials are: personal reports, statistical records, mass communication records, artifacts (objects made or used by people for a practical or artistic purpose), fossil or skeletal remains, or verbal reports by eye witnesses or informants. Since the social scientist did not supervise the production of these materials (as he does in observations, interviews, questionnaires, or experiments), he must have some way of determining whether the information contained in the material is accurate or whether the object is genuine. The procedures he uses to do this are called authentication procedures. These authentication procedures fall into two classifications: external and internal.

External Authentication

1. External authentication can be used for personal reports, mass communication records, verbal reports, artifacts, or remains.

2. It applies to the physical properties of the object itself, not to its meaning or content.

3. It attempts to establish the authenticity of the object or the accuracy of the report or record by external means:

    a. It establishes the date of 1) a document by chemical analysis of the ink or the paper or by determining when the writing implement was invented, 2) an artifact by chemical analysis of the earth surrounding it.

    b. It establishes authorship of a handwritten document by getting other samples of the person's handwriting.

    c. It establishes the physical ability of the person to report accurately (i.e., vision, proximity to event, ability to write or typewrite, education).

Internal Authentication

1. Internal authentication can be used for personal reports, mass communication records, verbal reports, but not for artifacts or remains.

2. It applies to the content of the records and reports, not to their physical properties.

3. It attempts to establish reasonableness of the content of the records and reports by internal means:

a. It reveals the motivations of the person making the record or report. Would he have any reason for lying, leaving out relevant information, or coloring his report?

b. It determines whether anything is out of its proper time context. Are there any anachronisms in the events, dates, customs, etc., that are mentioned in the reports or records?

POSSIBLE INSTRUCTIONAL ACTIVITIES

1. Read studies from the Resource Materials (for example, Cave Painting, Carbon 14), and after a discussion of types of authentication, have students orally state which types of authentication these studies use.
2. Analyze the data collected in class projects. After a discussion of mode, mean, median, and range, students could plot the data collected in these projects or in class quizzes. Students make graphs and distributions.
3. Have the students bring in studies from newspapers - graphs, etc. A display of the best examples could be put on the bulletin board, and a prize awarded for the most interesting example.
4. The class could take a field trip to a local concern which has converted to computers. The teacher should point out various types of computers, and emphasize the importance of computer analysis by quizzing the class when it returns to school.
5. Have students copy several graphs in their notebook, noting the value of neatness in scientific research.
6. The students could make up their own graphs, either from topics suggested by the teacher, or utilizing the data collected in class projects.
7. Using the studies in the Resource Materials (Cave Painting, Carbon 14, or Chimpanzee study), have the students analyze errors made in these studies. The teacher should first lecture on types of errors or identify them in a teacher-led discussion.

INTERPRETATION OF RESULTS

(SPECIFIC OBJECTIVES TWELVE AND THIRTEEN)

As you read through the research studies in the Resource Materials or in social science journals, you may doubt the validity of some of the studies. Perhaps you occasionally have a vague feeling that a conclusion does not seem logical or "right." Since social scientists are not infallible, there could be a research error or an unfounded statement in the study which casts real doubt on the interpretation of results. How can you clarify and test your doubts so that you can detect specific errors in research studies? The discussion about statements which are supported or contradicted by evidence and about more specific research errors should aid you in correctly identifying faulty research studies.

STATEMENTS SUPPORTED OR CONTRADICTED BY THE FACTS GIVEN

Deciding whether a statement is supported or contradicted by the facts given is a first step in distinguishing truth from falsehood or fact from opinion in a research study. It is easy to overlook and accept an unfounded statement or a conclusion drawn beyond the data if the statement or conclusions seem logical. Thus, very careful reading is necessary to differentiate statements which are contradicted by the facts from statements which are fully supported by the facts. Only the facts _given_ in a research study should be evaluated. Other sources of information would, therefore, be irrelevant.

RESEARCH ERRORS

Sampling error (Experimental study)

Sampling error occurs when a sample is not randomly selected and does not reflect adequately the larger population from which it is taken.

An example of a sampling error is one in which the researcher is too lazy to sample the _whole_ population of New York but tests only people in his own neighborhood. Chances are the sample would be _biased_ toward people who are much like the researcher in such factors as race and socioeconomic level.

Another example of a nonrepresentative sample is one which is unconsciously chosen by the researcher to support his hypothesis. Suppose the hypothesis is "Are high school girls smarter than high school boys?" If the researcher feels that high school girls are smarter, he may unintentionally choose most of the girls in his sample from the academically better high schools. The results of such an investigation, of course, would not be valid.

A general way to deal with representative samples is for the researcher to sample randomly from the population he is studying. Only by introducing unbiased representations into his sample is the researcher able to generalize his results.

Confounding variable (Experimental study)

A confounding variable operates in addition to the variable that is being tested in an experiment and distorts or hides the effect of the tested variable. A confounding variable, for example, confuses cause and effect relationships in an experiment.

For instance, if an experiment were set up to test the effects of gravity on a feather, a confounding variable might be wind. The wind would distort the effect of gravity so that the feather would float in the air for a long time before it finally landed on the ground. Unless the wind is controlled or eliminated, as in a vacuum, it would appear that gravity has much less effect on a feather than it actually has.

Authentication error

An authentication error occurs when either no attempt is made or inadequate attempts are made to verify historical or anthropological data.

When no attempt is made to verify data, a conclusion may be reached on the basis of false information. For instance, suppose a historian finds a typewritten letter that is dated 1776. The letter reveals valuable information about the political situation in England at the beginning of the American Revolution. If the historian draws any conclusions from the data in the letter, he will be wrong. He will have overlooked the fact that typewriters were not used until the nineteenth century.

Inadequate data for conclusion (Both experimental and nonexperimental studies)

A conclusion is based on inadequate data if it (a) goes beyond the information revealed in a study, or if it (b) describes a cause and effect relationship in a nonexperimental study.

As an example of drawing conclusions beyond the data, suppose a study is designed to reveal whether people who do much reading also go to the movies often. The researcher finds out that those who read more than four books a month go to the movies no more than twice a year. He concludes that people who read a great deal go to the movies very rarely and also go to very few parties. Since party-going was not tested in this study, it is obvious that the researcher has reached well beyond the information revealed in his study and has drawn an unfounded conclusion.

A cause and effect relationship can be shown only in an experiment. The following is an example of a conclusion which wrongly describes an associational relationship as a cause and effect relationship. Suppose the researcher concludes that reading frequently keeps people from going to the movies often. Nothing in the study showed that reading caused people to avoid movies. The study only showed that people who read a great deal also do not go to the movies frequently. Reading and attending movies are associated with, not caused by, each other.

POSSIBLE INSTRUCTIONAL ACTIVITIES

1. Using studies from the Resource Materials, teachers may omit some information from the studies and ask students to identify the omission.

2. Have various students report on research studies (from the Resource Materials or other sources) and have the class criticize the researcher's technique.

3. Have a few students volunteer to conduct small scale investigations and report on their difficulties in avoiding errors.

4. Lead a discussion of the importance of avoiding errors, such as confounding errors.

5. Take a trip to a nearby university to see the research facilities used in social sciences.

6. Give students practice in seeing if the researchers' conclusions are supported by evidence.

SECTION THREE: RESOURCE MATERIALS

RESOURCE MATERIALS

INTRODUCTION

The first section of the resource materials contains sixteen summaries of actual social science research studies reported in scholarly journals and popular magazines. Each of these studies is identified by discipline (and, in some cases, disciplines, because of the overlapping nature of areas of inquiry). A brief abstract indicates whether the hypothesis under investigation was causal or descriptive, the data collection procedure which was employed, and the type of analysis to which data were subjected, if any. In addition, for the first five studies each research component (sampling procedure, statement of results, etc.,) is identified and clearly labeled.* You may wish to make these studies available to your students, although the reading comprehension level demanded by them may be too high in some cases.

The next major section contains some additional material on data collection. Discussions and examples of projective tests, a treatment of confounding variables, and brief explanations and illustrations of basic research designs are provided.

Data analysis is the focus of the third and final section which includes supplemental description of statistical measures and examples of both external and internal authentication procedures for historical and anthropological inquiries.

You may, of course, use the information included in this section in any way you wish or ignore it completely. It is made available only to save you time and effort.

---

*For anthropological and historical studies primarily (but in other disciplines too) where the research considers artifacts, or extant records rather than people, the data collection process is not described but presumed to be physical examination.

Henry Clay Lindgren, "Age as a Variable in Aversion toward Food and Occupation," Journal of Consulting Psychology, vol. 26, no. 1, February 1962, pp. 101-2.

Abstract

To test his descriptive hypothesis that teen-agers have more dislikes than likes, Lindgren gave a questionnaire to college students over and under 21. His conclusion that teen-agers do have more dislikes than older people seems to be based on inadequate data. Teen-agers may have more dislikes than slightly older people because they don't have as wide a range of experience. But how do teen-agers compare with small children who are often very finicky about food or with elderly people who are said to be set in their ways?

Summary of Study

Descriptive Hypothesis

Are teen-agers really "mad at the world?" Do they really have more dislikes than likes? To establish the truth of these common descriptions of young people Lindgren arranged a study to test the hypothesis that people under 21 would reject more foods and occupations than those over 21.

Data Collection

He gave a questionnaire listing 39 undesirable or unfamiliar foods and 51 unpleasant jobs to a group of college students.

Subjects

The 392 college students were "asked to indicate which foods they disliked so much they would refuse to eat them, and which jobs (appropriate to sex) they would not do under any circumstances."

Data Analysis

The average number of disliked foods for men under 21 was 7.16, and for men over 21 the average was 4.41. Women under 21 disliked 6.72 foods, while those over 21 disliked 4.45 choices.

As for jobs, men under 21 disliked 5.70 and men over 21 disliked 4.08. Women under 21 did not like 6.61 jobs, while those over 21 did not like 4.88.

Statement of Results

Thus, it was shown that students under 21 show dislike for more foods and occupations than those over 21.

Interpretation of Results

Since rejection of both foods and jobs is so much stronger before 21 years of age, it would seem that young people in their teens do generally have more negative attitudes than they do when they are older.

POLITICAL SCIENCE AND SOCIOLOGY

Raymond E. Wolfinger, Barbara Kaye Wolfinger, Kenneth Prewett, and Sheilah Rosenhack, "America's Radical Right: Politics and Ideology."

Abstract

Wolfinger's study of the radical right fits rather neatly into the outline of the components of a research study described in the "Content Guidelines." To test his descriptive hypothesis (What is the education and income level, the religion, age, political part, attitude and motivation of the typical member of the radical right in America?), Wolfinger uses questionnaires, interviews, and statistical records.

Summary of Study

Descriptive Hypothesis and Subjects

To discover the nature of the radical right in American politics, Raymond E. Wolfinger and his research associates decided to investigate the individuals who attended and supported the 1962 Oakland, California "Anti-Communism School" of the Christian Anti-Communist Crusade.

Data Collection (Questionnaire-Interview)

From the data gathered from 214 mail questionnaires and 94 interviews, the researchers found that these people were usually highly educated and had high incomes. A greater proportion of the Crusade's supporters were Protestants than the general population of the area from which the Crusade drew its participants. Also, a disproportionate number of the crusaders were over 50 years of age.

Statement of Results

Politically, the Crusaders were almost unanimously Republican. Furthermore, at a time when a national sample of Republicans revealed that 46% of the party members supported Nixon and only 13% chose Goldwater, 58% of the Crusaders favored Goldwater. As for their attitudes toward Communism, the Crusaders were more likely than the average American to feel that American rather than foreign Communists were a greater danger. They particularly feared that American intellectuals have been infiltrated by Communists.

In discussing the resaons that Americans associate themselves with the radical right, the researchers considered alienation, status anxiety, provincialism, and fundamentalism or old-fashioned individualism.

The researchers discounted the common notion that members of the radical right are alienated--that is, isolated and cut off from American society. Rather, they found that Crusaders belong to many groups and are highly active politically. Further, they have strong confidence that their actions can accomplish things in the political arena.

The researchers did not find that the rightists were more worried about their social status than other groups. The sample had not moved more rapidly up or down the status ladder than the general population.

The Crusaders did not come from provincial (rural or small town) backgrounds any more than the average American.

Interpretation of Results

Wolfinger and his associates concluded that fundamentalism and old-fashioned individualism are the source of the radical right movements. The Crusaders fear the end of the old way of life and blame Communists for the displeasing changes in American society. The Crusaders flock to the Republican party not because the party is as radically conservative as they would like it to be but because it is the more conservative of the two political parties which have power in the United States.

As an explanation of the success of the radical right, Wolfinger suggests that the rightists become more active when compatible ideas are present in the mainstream of American politics and when the rewards for agitation are greater than the constraint.

Paul Ekman, Wallace Friesen and Daniel R. Lutzker, "Psychological Reactions to Infantry Basic Training," Journal of Consulting Psychology, vol. 26, no. 1, February 1962, pp. 103-4.

Abstract

Contrary to the common assumption, this study shows that military training does not have a beneficial psychological effect on recruits. Ekman and his associates reached their conclusion after conducting an experiment in which they used personality tests to collect their data. The experimental design, that is, the procedure used in the experiment, was the before-after design. The same group of recruits were tested repeatedly after they had completed one, four and eight weeks of basic training to see if the military training had an effect on their personalities.

Summary of Study

Causal Hypothesis--Data Collection

"One of the working assumptions of military psychiatry is that basic training has a beneficial psychological impact on the typical recruit. The possibility of this hypothesis was explored by administering the MMPI (a personality test of maladjustment) to recruits selected from the first, fourth, and eighth weeks of basic training..."

Sample

"93 active army personnel having 12 years of education and between 18 and 22 years of age served as subjects..."

Statement of Results

"There was no increase in scores on ego strength, or any other evidence of beneficial psychological effects accruing from basic training. The change in the shape of the profiles suggests that aggressive, impulsive, and energetic features became slightly more prominent."

Interpretation of Results

"The most cautious explanation of this change would be that the subjects were more willing to admit mild kinds of antisocial behaviro in the eighth week, since they had finished their training, and perhaps also experienced some relief at completing an arduous task. The changes on the sub-scales imply that more callous attitudes, a tendency to ignore the needs of others, and feelings of self-importance increase slightly during basic training. The recruits appear less prone to examine their own responsibility for conflicts, and more ready to react aggressively."

Franklyn N. Arnhoff, Henry V. Lion, and Irving Lorge, "Cross-Cultural Acceptance of Stereotypes Towards Aging," Journal of Social Psychology, vol. 63, June 1964, pp. 41-58.

Abstract

To test the common assumption that old people are less respected in America than elsewhere in the world, Arnhoff and his associates gave questionnaires to approximately equal numbers of college students in six countries around the world. The researchers made an attempt to select a representative sample of people in the world. But, three of the countries selected, the United States, Great Britain, and Sweden have Western European values. Furthermore, Japan and Puerto Rico are probably more influenced by Western European values than any other countries in Asia or Latin America. Another possible sampling error might arise because college students are not necessarily typical members of their societies. They are more exposed to foreign ideas than most nonstudents are and they are not usually as tied to their distinctive national traditions. It is possible that college students are more similar to each other than they are to their own countrymen.

Summary of Study

Descriptive Hypothesis

Do different nations hold different beliefs about aging? We have often heard that the United States is youth-oriented and that old people are less respected here than elsewhere. But studies that reveal that Americans have many negative opinions about old people do not show that Americans like old people less than other peoples do. Obviously, it is necessary to investigate the attitudes of people in other countries before comparisons can be made with American attitudes.

Data Collection (Questionnaire)

To determine the attitudes of different countries toward the aged, the subjects were given "The Attitudes Toward Old People Questionnaire" which is a 100 item list of statements about old people. The items were chosen from a longer list of statements collected by Tuckman and Lorge from writings which include misconceptions, stereotypes, and over-generalized beliefs about old people. In general the items are negative statements, i.e., they describe bad characteristics such as "are a burden to their children," are untidy and careless about their appearance," are grouchy," etc. The researchers justify the negative tone of the questionnaire by stating that if people are asked to make a list of characteristics associated with aging, most of the traits would be undesirable.

The questionnaire was translated into the various languages by bilingual nationals of each country who live in the United States. As a check the questionnaire was then translated back into English by other bilingual nationals. Thus the same ideas were communicated on the questionnaire no matter what language it was written in.

Sample

The subjects for the study were between 423 and 184 college students from 6 different countries. College students were used to make the subjects "as equivalent as possible without undertaking a major field study." The countries chosen were the United States, Great Britain, Sweden, Japan, Greece and Puerto Rico. These nations represent differences in economic status, family structure, degree of industrialization, and social welfare programs for the aged.

Statement of Results

Most items produced significantly different responses from the different groups of students. But the differences are often quite small and meaningless. To determine the overall agreement of each country to the statements on the questionnaire, an overall score was obtained from the individual scores for each item.

Scores for Overall Agreement with Stereotype

	United States	Sweden	England	Puerto Rico	Japan	Greece
Average % agreement	46.0	49.4	52.2	56.6	57.3	66.5

Statistically significant differences in attitude were found only between Greece and the United States, Sweden and England.

Interpretation of Results

This study could provide data for further studies on the role of social, economic, psychological and political factors in the establishment of attitudes toward the aged and on the importance in the functioning of the aged in different societies.

The study showed that aging in a variety of countries is accompanied by many stereotyped beliefs which are largely negative. Despite differences in geography, education, and culture, in the different countries, the attitudes toward old people were very similar.

Could it be that the physical and mental changes connected with aging are brought about by people fulfilling their own expectations?

> Societies tend to be age graded, with expectations and appropriate roles determined by many years of tradition and societal function, and far less by individual assessment and actual capabilities. The roles, statuses, and physsical and psychological characteristics of different stages in the life cycle are soon learned and incorporated into one's belief system... Thus the wide acceptance of these beliefs about the concomitants of aging cannot be taken as some sort of proof that they actually do occur, but rather that these beliefs themselves play a role in setting the stage for the person to behave in the manner expected...

The study is important in putting America's attitudes toward old people in the proper perspective. It has been assumed that America is unique in its youth-orientation, negative attitudes toward the aged, tendency to throw away or trade in old things, and materialism. But contrary to these common assumptions, America is not only not unique in these attitudes, but is less negative about old people than other nationalities are. The aged always have been and will continue to be a problem for other members of society no matter where they live.

Richard Panek and Thomas Hannum, "Relation between Autokinesis and Introversion/Extraversion," Journal of Consulting Psychology, vol. 26, no. 5, October 1962, p. 477.

Abstract

Panek and Hannum set out to test the descriptive hypothesis, "If people see apparent movement in a stationary object, will they be likely to be outgoing or withdrawn?" They use personality tests to determine whether the subjects are outgoing or withdrawn.

Summary of Study

Autokinesis = apparent movement of a stationary object.

Extraversion = personality trait of liking to be with other people and of being concerned with objective reality.

Introversion = personality trait of preferring to be alone and of living in an inner world.

Descriptive Hypothesis

If people have a tendency to see apparent movement in a stationary object (such as a pinpoint of light in a dark room) will they be likely to be outgoing or withdrawn?

Subjects

To answer this question Ranek and Hannum chose 25 male and 25 female subjects who were volunteers from a freshman psychology course.

Data Collection

The students were given a number of personality tests to see whether they were outgoing or withdrawn, merry or depressed, or whether they have extreme shifts in mood.

Statement of Results

The hypothesized relationship between seeing the stationary light move and sociable or retiring personalities was not found. But some rather unexpected results were obtained.

Seeing stationary objects move seems to be related to maladjustment more than to outgoing or withdrawn personalities. The more likely people are to see unmoving lights move, the more likely they are to be depressed, emotionally unstable, or otherwise maladjusted. Males tend to see more false movement than women and also score higher on maladjustment tests. But even women who tend to see the stationary light move also tend to be maladjusted.

Robert J. Rodden, "An Early Neolithic Village in Greece," Scientific American, April 1965, p. 83.

Abstract

Rodden's study, based on a descriptive hypothesis, of the northern Greek village of Nea Nikomedia reveals that Europe entered the agricultural stage of development much earlier than it was previously thought. Carbon-14 dating has shown that the agricultural village existed around 6220 B.C. By comparing artifacts found at Nea Nikomedia with those from later sites in Europe and in the Near and Middle East, Rodden reached a new conclusion that Europe did not borrow all of its early culture from move developed areas in the East, but made some unique contributions to Eastern development. Both Carbon-14 dating and comparing the physical characteristics of artifacts are examples of external authentication.

Summary of Study

Eight thousand years ago the Macedonian plain of Northern Greece was flooded by a bay or a shallow lake. Along the edge of the water, farmer-herdsmen raised wheat, barley, and lentils and tended sheep, goats, and possibly cattle and pigs. This information doesn't seem particularly startling until one realizes that it has long been accepted that Europeans did not enter the agricultural stage of development until a much later time. The discovery of the village of Nea Nikomedia, the oldest Neolithic community yet found in Europe, reveals that the agricultural revolution of the Neolithic period took place in Europe at a very early time.

Samples of organic material from the site were sent to the Radiocarbon Dating Laboratory at the University of Cambridge where they were dated as being from 6220 B.C. plus or minus 150 years.

Excavating Nea Nikomedia to find out how the people lived was especially difficult because the ancient people built their homes with mud walls on wooden frameworks. The clay of the earth and of the walls have blended now so that any details of ancient buildings would appear to be only faint discolorations in the surrounding earth. To define the outlines of the houses, the "painstaking technique of scraping" had to be used. Soil had to be removed from the "possible" walls an inch at a time. The site is a knoll that contains 2 ancient villages one on top of the other. The older one was circled by walls on the landward side. The newer one was surrounded by a water-filled ditch that may have been a protective moat. In contrast to Neolithic villages in the Middle East and Asia Minor in which the houses were attached to each other around a courtyard, Nea Nikomedia's people built their houses several yards apart. It is assumed that the houses had peaked and thatched roofs to carry off rainwater. Since the largest room in the older village contained five figurines of women, it is possible that it was the central place of worship or ritual homage to fertility.

The bones of pigs and cattle found in the village indicate that these animals may have been domesticated. If so, 6000 B.C. would be "earliest dated occurrence of domesticated cattle yet known anywhere in the world." This find suggests that domestication of cattle may have originated in South East Europe and in Asia Minor where early cattle bones were also found. The ancient Greeks may have domesticated pigs independently of the ancient people of Iraq who also kept pigs.

Most probably the agriculture and the raising of sheep and goats came to Greece from the Middle East. Many of the tools and weapons found in Nea Nikomedia are typical of other Neolithic communities in Europe or the Middle East. But certain artifacts, such as those made of chipped stone, are closer to those from Central Greece, Yugoslavia and Bulgaria than from Southern Greece, Asia Minor, and the Middle East.

The pottery was well-made and decorated with finger impressions or geometric patterns of red on a cream-colored background. It is similar to the red and cream pottery found in later settlements in Central Greece and several hundred years later in Asia Minor. Also human faces with pinched-up noses and applied oval clay eyes appear in Nea Nikomedia and later in Asia Minor and Iraq.

The belt fasteners found in Nea Nikomedia resemble those from other parts of Europe. But, neatly carved and polished marble and serpentine hair and ear jewelry look like similar articles from the Middle East.

The clay sculptures of men and women are particularly artistic. They are stylized forms with long narrow heads, large noses and fixed poses. Figurines of sheep and goats were rather crudely made, but three beautifully carved and polished blue and green serpentine frogs were found. "...what significance the frog may have possessed that inspired the execution of its portrait is unknown."

The people of Nea Nikomedia seemed to care little about their dead. Graves do not seem to have been specially prepared. In some cases people seem to have been stuffed into tiny holes in the ground. No personal adornment or food has been found in the graves.

The study of Nea Nikomedia shows that the area had an established Neolithic civilization that developed a culture and a building style suitable specifically to European conditions. The uniqueness of the houses and artifacts in ancient Greece indicate that the poeple of the Middle East and Asia Minor were not alone in their contributions to Old World Civilization but were aided by the early Europeans.

James E. Hill, Jr., "El Chamizal; A Century-Old Boundary Dispute," *Geographical Review*, October 1965, p. 510

Abstract

In writing his report on the boundary dispute, Hill consulted the various treaties which dealth with the issue, the report of the 1910 Commission, and other *official records*. He also used *statistical records* to find out *what* and *how much* property would be involved in the land exchange as well as how much tax money El Paso would lose in the transaction. A *descriptive hypothesis* guided this.

Summary of Study

Since the Rio Grande was made the boundary between the United States and Mexico in 1848, the river has changed its course several times. In most places the river has moved in remote areas where it caused no political problems. "But in the El Paso-Ciudad Juarez area the southward shift of the river had, by the end of the 19th century, added about 598 acres to the north bank." For 100 years this area, called El Chamizal, had repeatedly provoked political arguments. Finally, in 1962, President Kennedy and President Adolfo Lopez Mateos settled the boundary question and paved the way for the American-Mexican Chamizal Convention Act of 1964.

One reason for the long unsettled boundary question was the ambiguous wording of the various treaties defining the boundary, especially that of 1884 which tried to anticipate problems associated with the wandering river. The treaty of 1884 stated that the center of the river would continue to be the boundary if the river moved by normal slow and gradual erosion. But if a forceful current created a new river bed or two river beds, the line should stay along the original river route, whether wet or dry. In 1895 Mexico claimed that the river had shifted rapidly, but the United States argued that the change was normal and gradual. Later, to reduce flooding the northward curve of the river was cut off, leaving on the American side a Mexican area called Cordova Island. Cordova Island created a very annoying problem for both nations. It cut El Paso in half but could not be developed by Mexico because of the neck of land which connected with the rest of Mexico, which was only wide enough for a single road.

In 1910 the United States and Mexico agreed to settle the issue through a Commission consisting of an American, a Mexican, and a Canadian arbitrator. The Commission voted that part of the El Chamizal should belong to the United States and part to Mexico. But the United States refused to accept the decision because the Commission was not authorized to split the territory but only to award it to one side or the other. President Kennedy reversed the American policy on the border question and decided that the United States should accept the 1910 Commission's decision.

The 1964 American-Mexican Chamizal Convention Act decided that the river would be rerouted along a concrete course determined by engineers and that the boundary would follow the river. The land Mexico ceded to the United States was vacant and owned by the Mexican government, but the United States territory ceded had to be freed of 5,600 people, buildings, communication lines, etc., bought by the United States government, then given to Mexico. Most of the ceded American territory is industrial and agricultural (railroad and stockyards).

The annual tax loss to El Paso from the land exchange will be $201,591 plus another $306,895 which would probably have come from areas the University of Texas had planned to sell. Even though the United States appears to be losing more than Mexico in terms of improved land and taxes, the exchange can benefit Americans in the long run.

"The Convention of 1964 is one of the most important diplomatic steps the United States has taken toward increasing good relations with Mexico and other Latin American nations. The long-term psychological effect of the Northern Giant's ceding valuable urban-industrial land to a less well developed Latin-American neighbor is worth far more than any soon-forgotten financial aid. In addition, had we decided to prolong the dispute indefinitely, we should only have continued to stifle the growth of El Paso... Had Mexico insisted that all the land to be ceded to her be in El Chamizal, as she did in the past, the United States would have continued to reject the proposal. With each nation respecting the boundary problems of the other and negotiating with an attitude of friendship, cooperation, and mutual understanding, a century-old boundary dispute has come to an end."

The 1964 American-Mexican Chamizal Convention Act decided that the river would be rerouted along a concrete course determined by engineers and that the boundary would follow the river. The land Mexico ceded to the United States was vacant and owned by the Mexican government, but the United States territory ceded had to be freed of 5,600 people, buildings, communication lines, etc., bought by the United States government, then given to Mexico. Most of the ceded American territory is industrial and agricultural (railroad and stockyards).

The annual tax loss to El Paso from the land exchange will be $201,591 plus another $306,895 which would probably have come from areas the University of Texas had planned to sell. Even though the United States appears to be losing more than Mexico in terms of improved land and taxes, the exchange can benefit Americans in the long run.

"The Convention of 1964 is one of the most important diplomatic steps the United States has taken toward increasing good relations with Mexico and other Latin American nations. The long-term psychological effect of the Northern Giant's ceding valuable urban-industrial land to a less well developed Latin-American neighbor is worth far more than any soon-forgotten financial aid. In addition, had we decided to prolong the dispute indefinitely, we should only have continued to stifle the growth of El Paso... Had Mexico insisted that all the land to be ceded to her be in El Chamizal, as she did in the past, the United States would have continued to reject the proposal. With each nation respecting the boundary problems of the other and negotiating with an attitude of friendship, cooperation, and mutual understanding, a century-old boundary dispute has come to an end."

M. I. Artamonov, "Frozen Tombs of the Scythians," Scientific American, May 1965, p. 101.

Abstract

External authentication by Carbon-14 dating was necessary to support the descriptive hypothesis and to establish the dates of the frozen Scythian tombs found in Siberia. The wealth of data about the way of life of the ancient Scythians neither supports nor contradicts the conclusion reached by the author of the article that art has been important to men in all times and among all people and that it shapes ideological concepts. It could very well be that ideologies shape art. At any rate, there seems to be inadequate data for the conclusion reached by Artamonov.

Summary of Study

Our view of the everyday life of ancient people has been distorted by the fact that organic materials such as wood, leather, cloth and so forth disappear and leave only stone, clay, bone, or metal objects as records of ancient life. In places of permanent wetness (Swiss peat bogs) or dryness (Egyptian deserts) a fuller picture of early peoples is left. In the last years, Soviet archeologists have discovered another means of preserving ancient material--freezing. In a number of burial mounds in the Altai Mountains between Siberia and Outer Mongolia, archeologists have found the refrigerated bodies of 2,000 year old chieftans, their horses, clothing and possessions. The tombs contain people who the ancient Greeks and Iranians called Scythians--nomads who road across the Eurasian plains. Herodotus and other Greek writers went so far as to describe the wandering Scythians as "fabulous baldheaded creatures: the Argippaei, the one-eyed Arimaspi and the gold-guarding griffins... (half lion and half eagle)." The Eurasian Scythians spread from the northern frontiers of Greece, Iran, and China and traded with the absorbed aspects of all of these cultures. But despite their contacts with other peoples, they maintained a distinctive way of life and pastoral economy.

The Altai Mountain tombs have stripped away much of the mystery and mythology that has distorted the true nature of the Scythians. Most of the entombed bodies have been found to be Caucasian Europeans, not baldheaded or one-eyed monsters. Since many of the objects found among the buried men are made in Iranian style, the author concludes that the people of the Altai Mountains were "clearly of Iranian origin." However, the Scythian nomads had close contact with Asian peoples as well as Western groups. Some of the people in the tombs were definitely Mongoloid. Furthermore, some of the objects were of Chinese origin.

The dating of the mounds is not exact; judging from the tree-growth rings in the timber, the graves were built over a period of 200 years. Carbon-14 dating is accurate only plus or minus 130 years. Various authorities estimate the ages of the mounds between 700 B.C. and 100 A.D. Since the culture of the Scythians must have taken some centuries to spread from Iran, it is unlikely that the graves could have been made before 500 B.C.

All of the tombs found by the Soviet archeologists had been plundered by treasure hunters many centuries ago. Clues such as separate layers of clean (old) and dirty (newer) ice indicate that the grave robbers arrived several generations after the relatives of the dead men could prevent their pillaging. But the timbers protecting the graves were cut with primitive metal axes; thus the pillagers were also ancient people--probably Turks who invaded the Altai Mountains around the third century B.C.

Even though the graves were robbed of much of their original contents--particularly jewelry--much has remained to reveal the way of life of the Scythians. To build their tombs the Scythians dug pits 25 feet square and 15 feet deep. A wooden framework was built in the center of the pit for the chief. Outside of the wooden walls, 6-22 slaughtered horses were buried. The inner structure was a small room with a board floor and felt-covered walls and ceiling. The body of the chieftain, sometimes with his wife, was placed in a large coffin which was covered with carved figures of animals or designs. Funeral goods such as tables, stools, and dishes with food were also in the tomb. The chamber was covered with logs, bark, soil and boulders to form a mound as much as fourteen feet above the ground. The cold air which gets trapped in the insulated freezing chamber cannot warm up enough in the short summers to allow the organic matter to decay.

The bodies of the dead were carefully embalmed. The brains and internal organs were removed and the bodies were stuffed with grass or hair so that their shapes would be maintained. One man was covered with intricately beautiful tatooed designs of animals. The people's clothing was made of leather, fur, felt, or woven hemp.

The men wore narrow leather trousers, felt stockings, high boots with soft soles, a cape-like tunic with long decorated sleeves, and a felt hat with leather earflaps. The women wore similar tunics and leopard fur boots with elaborate embroidery on the soles. Perhaps they sat crosslegged to show the soles of their shoes. The women wore varied and elaborate headgear and hairstyles.

Most of the artifacts found in the Altai tombs were decorated in the unique Scythian "animal style"--graceful and stylized animal forms. What remains of the jewelry shows that the ancient nomads made beautiful adornments of gold, bronze, silver, turquoise, and horn. An incredibly complexly woven rug found in one of the graves is the oldest known woolen article.

Because the grave robbers did not bother sorting through the dead horses to gather loot, the horse burials have shown the best picture of the artistic talents of the Scythians. Horses were vital to the nomads' life and received careful attention. Their manes were clipped and their tails were braided. The saddles and saddlecloths were decorated with fantastically elaborate designs. Felt, leather, colored thread, gold inlay, and intricately carved and gilded pendants covered the trappings. Apparently, the horses wore beautifully decorated masks over their heads for everyday purposes.

The Altai remains show that art has been important to men in all times and among all peoples. "In every culture and age art has served not only to fulfill his aesthetic needs but also to shape his ideological concepts within the framework of his social relations."

Harold M. Rose, "The All-Negro Town: Its Evolution and Function," Geographical Review, July 1965, p. 362.

Abstract

Geography, sociology, and history are very much intertwined in this study of the all-Negro towns. The overall discussion of the locations and the development of the towns fits within the study of geography, but the discussion of American Suburbs fits neatly within the jurisdiction of sociology, the discussion of the origin and development of all-Negro towns in America required historical research. Statistical records helped the researcher to collect his population of towns, to rate the economic status of their residents and to gather data in support of a descriptive hypothesis.

Summary of Study

Few people who do not live in or near all-Negro towns in the United States realize that they exist. All-Negro towns are defined by Mr. Rose as "all places with a population of 1,000 or more of whom more than 95% are classified as nonwhites." Most of the 12 such towns "which are physically or politically separated from their nearest neighbors" have a white population of less than one percent which does not dominate the community. Another seven towns with Negro populations over 95% are classified as "pseudo towns" because they are not politically, economically, or physically separate from larger cities. Ten of the 12 all-Negro towns are politically independent cities with a mayor-council form of government. This form of government is most prevalent in the United States as a whole, especially in working-class communities.

The pseudo towns are listed as separate places in the 1960 census, but are actually branches of a larger community in which they are not eligible to vote. All of these towns are in the South. They were formed because Negro communities developed on the fringe of larger towns or because the larger town expanded to reach their limits. The seven pseudo towns are Samtown (Alexandria), Louisiana, Greenwood (Tuskegee), Alabama, Gifford (Vero Beach), Florida, Pleasant Hill, Saratoga Place and Lloyd Place (Suffold), Virginia, and Bennettsville, Southwest (Bennettsville), South Carolina. None of the pseudo towns have populations over 5,000.

All-Negro towns emerged in four fairly distinct periods: pre-Civil War, post-Civil War, during and after World War I, and post-World War II. The first all-Negro towns were not in the South. These early towns were established near Northern towns by people who supported the Underground Railroad and the abolition movement. Hostility toward the Negro and the fear of being swamped by large numbers of escaped slaves caused the Negro communities to be isolated. The first all-Negro town in the pre-Civil War days was Brooklyn, Illinois, near St. Louis. Later, Lawnside New Jersey near Philadelphia was established.

After the Civil War, Mound Bayou, Mississippi, Kinloch, Missouri, Urbancrest, Ohio, and Grambling, Louisiana, were formed. Mound Bayou was founded by the former slave and body servant of Jefferson Davis, Isiah T. Montgomery. Gambling grew up around a sawmill. Kinloch and Urbancrest started as agricultural villages near large towns.

World War I slowed European emigration and allowed Southern Negroes to move North to claim jobs. The agricultural slump in the South encouraged them to leave. Robbins, Illinois, Lincoln Heights, Ohio, Fairmount Heights, Maryland, and Glenarden, Maryland, were formed. Robbins was founded in 1917 as an all-Negro town by Eugene S. Robbins, a Negro realtor. In the early days it was a village of homemade tarpaper shacks. Lincoln Heights was organized by a lumber company in 1926. Today it is the largest and one of the most stable of the all-Negro towns.

Since World War II, North Shreveport, Louisiana, and Richmond Heights, Florida (near Miami) have evolved. They were built by housing developers for Negro buyers. North Shreveport was established on the foundation of a tiny Negro community by housing developers who relocated Negroes displaced by a freeway. Richmond Heights needed more and better housing in the Miami area.

Although all but two of the all-Negro towns are close to metropolitan areas, they are not really suburbs. Suburbs have not been clearly defined by sociologists, but the concept implies an ideal middle class settlement of central city emigrants who continue to work in the city. Like other suburbs, all of the metropolitan all-Negro towns are dormitory towns from which the residents commute to the city to work. But since most of the workers are blue collar manual laborers or union members, the all-Negro towns are not like the usual suburbs, but are "working class" suburbs. The all-Negro towns are generally distinguished from their nearest neighbor by a much higher percentage of substandard housing. The all-Negro towns are more similar to central city all-Negro areas than they are to their neighbors. Thus it seems that all-Negro towns are not like the more truly suburban towns around them.

Neither Mound Bayou nor Grambling could be called suburbs. The rural town of Mound Bayou is a farming center in the South's largest cotton producing country. Grambling is a college town with a relatively high income level.

In the future, some of the all-Negro towns may become more like suburbs when higher quality housing is built for middle class central city immigrants. But other towns are doomed to remain poor and crowded non-suburbs because of the poverty of Negroes in the area and opposition to all-Negro communities. The future of all-Negro towns depends on many unpredictable variables: Community attachment, barriers to Negro settlement in other suburbs, economic status of central city Negroes, philosophy of militant groups, and a tax base to support local development.

This overview of the all-Negro towns in the United States suggests that Negroes have been so isolated from American life that they have been able to maintain all-Negro communities that are not "threatened" by white immigration.

John Marco Allegro, "The Untold Story of the Dead Sea Scrolls, Harper's, August 1966, pp. 46-54.

Abstract

Allegro uses internal authentication to determine whether the New Testament as it is known today is an accurate translation of the original documents--a descriptive hypothesis. He finds it more reasonable to assume that the original sources were written with a liberal use of puns, words which could have several meanings. He suggests that many of the names or titles given to Biblical characters in the Greek and subsequent translations should have been translated to refer to religious titles of members of the Essene sect who wrote the Dead Sea Scrolls.

Summary of Study

"The first discovery (of the Dead Sea Scrolls) was quite accidental. Muhammad, a Bedouin goat herder, on the trail of one of his flock, chanced on a cave set high in the cliffs bordering the northwest shore of the Dead Sea in Jordan. He found inside some tall, wide-necked jars. In them were parchment scrolls wrapped around with some evil-smelling linen rags. These were the first of the Dead Sea Scrolls, seven in all, and now no money on earth would buy them. They included manuscripts of parts of the Old Testament a thousand years older than anything hitherto seen. Muhammad raised perhaps $60 on their sale to a Bethlehem cobbler. Four were smuggled to the United States and sold for a quarter of a million dollars; the other three found their way more directly to the Hebrew University in New Jerusalem, where all seven now attract thousands of visitors a year to Israel's Shrine of the Book.

"Since then Muhammad and his friends have been Scroll-hunting in earnest, well aware of the monetary value of every scrap of written material they can fist out of the dust. Another ten caves containing documents were discovered in the next nine years in the vicinity of the first find. They have produced the remains on parchment and papyrus, mostly fragmentary, of about four hundred different documents, a third of them texts of the Old Testament. All these scrolls have remained in Jordan. From 1952 the Bedouin began looking farther afield, so that today we have to acknowledge the fact that the whole of the west coast of the Dead Sea is a veritable treasure-house of ancient manuscripts...

"There (has been) no doubt that in the kind of Judaism exhibited by the Scrolls we (have) the religious matrix of Christianity... Scholars had long ago suggested that a Jewish sect called the Essenes might prove to be the missing link between normative Judaism and Christianity. Hitherto, we had known of these people only through the works of the ancient historians, like Josephus and Philo of Alexandria, writing in the first century of our era. It was generally recognized that in the Scrolls we had the remains of a vast Essene library. Furthermore, among the fragments were found traces of books we had known previously in later translations... So with caution these too, could be used to supplement our knowledge of Essene ways and thought.

"The Essenes were noted for their extreme piety. They lived in communal settlements loosely attached to towns and villages throughout Judaea but as far as possible keeping themselves apart. They had a mother community by the shores of the Dead Sea, and most scholars quickly identified this with the ruined monastery of Qumran (where some of the Dead Sea Scrolls were found). The Essenes were great Bible readers, and sought in every word of Scripture a message for their own day and age. They practiced baptism, and a form of communism, sharing their worldly wealth and caring for the sick and aged from a common fund. They sought in natural phenomena "signs of the times" and believed they could foretell the future. They had the powers of healing, combining esoteric knowledge of medicinal herbs with power over the demons of the spirit world.

"With all this newly discovered pre-Christian literature at hand we could now see the New Testament in clearer historical perspective than ever before. Certain aspects of Christian teaching merged satisfactorily into its sectarian Jewish background. But there were differences (that lead some scholars to believe that the early Christians and the Essenes were different people)...

"We are, I am sure, on the verge of a tremendous breakthrough, and it will start with a recognition of the full extent of Christianity's debt to Essenism. We might begin with a fresh examination of the names and titles of Jesus and the Apostles. If these can be shown to be specifically Essene in meaning and origin, then we have the kind of concrete link with the people of the Scrolls that goes beyond mere community of religious outlook.

"Something of this sort has long been recognized. Thus the paqid ('Inspector, Overseer') of the Qumran community was early on linked with the synonymous episkopos ('Bishop') of the Church. The 'Many,' as designating the generality of believers in the book of Acts, was seen to represent the Hebrew rabbim of the Scrolls. We can now go much further than this.

"A newly deciphered document refers to one administrative official by a Semitic word which must underlie the nickname Chephad given to Simon Peter. The Essenes clearly deemed it a rather 'special' word, since it signifies one having the ability to read men's minds through their faces. This gives us the clue to the origin and purpose of the story in Matthew 16 where it is Peter who recognized the Messianic calling of Jesus. Furthermore, since Peter is here and elsewhere being designated an 'Inspector, Overseer' on a pattern with the Essene administrative functionary, we can now see that many of the other stories related about him, speaking with tongues, relating the wonderful works of God, supervising the admittance of new members into the community, handling the common fund, and so on, are simply demonstrations of the supervisory work required of the Essene administrator...

"When we come to examine afresh the New Testament descriptions of some of Jesus' more unorthodox table companions, the 'gluttons' and 'wine-bibbers,' the 'harlots' and 'publicans,' we find that they too disguise Essene titles and self-descriptions. Of particular interest in this respect are the female 'sinners'... The Semitic word from which this word is derived means also 'angel, one who serves God' and reappears not only similarly disguised in what Josephus tells us of the Essenes but straightforwardly as the title of Jewish and Judaeo-Christian sects elsewhere in the Mediterranean world... These and many similar instances of word-play involving important religious titles and self-descriptions must prompt us to ask how these neo-Essene writers of the New Testament could bring themselves so to maltreat their source material... Quite obviously no female Essene would have called herself a 'harlot' any more than her male companion would have perverted his most coveted title of 'Chosen One of God' into 'tax collector, publican.'... We can only assume that in the Christian writings we have moved out from the central core of Essenism into a shadowy half-world where even the most sacred names and ideas of the original traditions can be changed to suit the storytellers' purpose...

"The process of unlocking these secrets begins with an attempt to find certain key Aramaic (the Semitic language much of the New Testament was written in) words and phrases represented by the Greek (the language into which the Aramaic texts were translated). Each word will be capable of a number of different interpretations and we shall expect to find it thus variously used in diverse parts of the Gospels and Acts. Where this proves to be the case we can cross-check our supposed Semitic original so that the accuracy of our choice is self-demonstrating. We have then to decide which of the possible meanings must be deemed basic to the storytellers' purpose, and we shall usually be right if we assume it to be the one that does not rise to the surface at all. In other words, the saying and incidents that appear in their open Greek form are of less importance, or indeed of none at all, to the writers' real purpose.

"This intricate process of juggling with words is not easy to understand in terms of Western thought and language. The Semitic family of tongues, however, lends itself readily to such punning. In ancient writing only the consonants were shown, the vowels being of less importance. For example, the group D*B*R* can mean 'word' or 'he spoke' or 'plague' or 'pasture,' depending on the context, the reader supplying the appropriate vowels...

"To the Jew of ancient times such playing with words was by no means a low form of wit. Similarities between words of this nature could be expected to have real significance, particularly where they occur in Holy Writ. So to deduce a teaching or to portray an incident quite different from the plain meaning of the Bible text was considered...(legitimate)...

"Certainly we are now only at the beginning of a complete revolution in our appraisal of New Testament traditions and their purpose. But already it is clear that there is scarcely a word of the Gospels and Acts that can be taken at its face value. The dead Sea Scrolls together with historians' records of the Essenes and the wealth of intertestamental literature already in our possession have at last given us the key with which to open the hidden mysteries of the New Testament."

Arend Lijphart, "The Analysis of Bloc Voting in the General Assembly: A Critique and a Proposal," American Political Science Review, vol. 57, 1963, p. 902.

Abstract

Lijphart's careful use of the statistical records to support a descriptive hypothesis of United Nations voting reveals that the commonly recognized blocs are not always voting blocs. That is, nations with regional or organizational associations which are thought to form voting blocs in the United Nations do not necessarily vote alike. He suggests that blocs can best be defined, identified and measured in terms of their actual voting records.

Summary of Study

Since 1950, new nations which are not bound firmly to the Western and Eastern blocs have flooded into the United Nations and have formed new "blocs." These new blocs are extremely important because the bloc members have the power to control the United Nations if they can control the General Assembly. The General Assembly has become the hub of the United Nations since the Security Council became paralyzed by Great Power disagreements.

Everyone agrees that it is important to study bloc behavior in order to predict and control international conditions, but no one agrees on the definition of a bloc, that is, the identification of bloc members or the measurement of cohesiveness necessary for a bloc's existence. Definitions range from regional groups, to political organizations, to voting behavior in the United Nations. Most discussions of blocs point to the Afro-Asian or Arab-Asian, the Arab, the Latin American and the Soviet blocs. Beyond these few groups, the identification of blocs becomes confused.

Thomas Hovet, Jr. improved the definition of blocs when he classified them into 5 groups:

1. caucusing blocs with bloc discipline--
   For instance, all members of the Soviet bloc must vote alike.

2. caucusing groups without bloc discipline--
   The Afro-Asian, Arab, African, Benelux, Commonwealth, Latin American, Scandinavian, and Western European blocs often vote similarly within their groups, but they are not compelled to be and rarely are in complete agreement.

3. geographic distribution groups

4. common interest groups--
   The colonial powers and Moslem nations often vote together on issues that affect their interest.

5. temporary groups

However, Hovet's identification of blocs does not provide a measure of cohesiveness or bloc-like behavior. In other words, he does not show the degree to which bloc members consistently vote together. Lijphart suggests that a measure of cohesiveness need not determine reasons for voting similarities (as some measures attempt to do). He favors the Rice-Beyle system of identifying blocs and measuring cohesion simultaneously by analyzing voting records. This system defines a bloc as a group of nations in which the voting agreement of all possible pairs of nations is equal to or greater than an assigned minimum figure. Lijphart modified the Rice-Beyle system by including abstentions in the tabulations and by ignoring absences rather than counting them as NO votes.

To apply this method of identification and measuring the cohesiveness of blocs, Lijphart studied the alignments on colonial issues in the General Assembly between 1956 and 1958. The issues were classified as colonial by subjective analysis. Objective procedures determined the choice of the important colonial issues. The important issues were those which were not unanimous and which were included in the United Nations' Yearbook.

Lijphart's analysis reveals a surprising pattern of blocs which do not fit most of the usual classifications. For instance, if countries that vote together 95.5% of the time are considered to be blocs, only four blocs exist in the world: 1) Soviet Union, Eastern European, 2) Iraq, Saudi Arabia, Jordan, and Libya, 3) Denmark, Norway and Sweden, 4) Western European and the English speaking nations of the Commonwealth.

At the 87.5% level the blocs are rearranged:

1. Soviet Union, Eastern European, and many African and Asian nations
2. Ceylon, Ethiopia, Tunisia, Greece and Yogoslavia
3. Scandinavia, Western Europe, Commonwealth, Portugal and Italy

At the 75% level, the blocs present still another configuration:

1. Soviet Union, Eastern Europe, Asia, Africa
2. Western Europe
3. Japan, Phillipines, Thailand, Pakistan
4. Latin America
5. Latin America
6. Latin America

The blocs revealed in Lijphart's study differ considerably from those found by Hovet. The Soviet Caucusing Bloc is also a voting bloc. Of the Afro-Asian group, 8 are not in the Asian-African-Soviet voting bloc. The Arab Caucusing group has high voting cohesion. The African Caucusing group isn't a clear voting bloc, but it includes nations with various degrees of voting cohesion. The Commonwealth Caucusing group is badly divided on colonial issues. Benelux and the Western European Caucusing groups are not separate blocs. The Scandinavian Caucusing group is too close to other European countries to be considered a separate bloc. The Latin American Caucusing group is actually 3 voting blocs. Greece and Yugoslavia vote much like the Asia-Africa-Soviet bloc. Austria, Finland, the United States, and Portugal vote with Europe and the Commonwealth. China, Ireland, Israel and Spain are unaligned at the 75% voting cohesion mark.

> In short, the actual pattern of bloc voting on colonial questions during the 1956-1958 period were sufficiently different from the pattern of caucusing and other formal groups to warrant the conclusion that the analysis of bloc voting on the basis of predetermined groups is unreliable.

N. J. Van Warmelo, "A Tale of a Heap of Stones," Africana Notes and News, vol. 16, no. 7, September 1965, p. 2.

Abstract

In an effort to locate some of the storied South African landmarks and collect evidence to support his descriptive hypothesis isi Vivane, Van Warmelo consulted African dictionaries, the personal report of another anthropologist, and interviewed South Africans who might be able to direct him to the sites of the heaps of stones.

Summary of Study

Travelers in South Africa have long told about the heaps of stones that appeared along recognized routes and were regarded by the Natives with "superstitious awe or respect." In both the Xhosa and Zulu languages they are called isi Vivane. But they are said to mean different things in the two languages.

In Xhosa: "A heap of stones thrown together by travelers at certain steep and dangerous passes on a difficult, tiring, journey, a small stone being added by every passerby, who says...'God help me...or ...give us strength.'"

In Zulu: "Cairn, accumulated heap of stones, memorial collection of stones, 'luck heap' (on which Natives throw stones as they pass at crossroads or some outstanding place, the action being believed to bring good luck)."

Van Warmelo describes his search for the heaps of stones as difficult: "For sure, everybody knew what an isi Vivane was. No trouble at all. But we wanted to see one. Well, that apparently was another matter. Eventually one reliable official was found who said he knew it since his earliest days as a herdboy... We searched around without success...only a few months earlier the Roads Department had scooped the whole heap up and carted it away for gravelling a road somewhere else. One more relic of ancient times destroyed!"

Quoting Dr. D. W. Hammond Tooke, Van Warmelo illustrates how anthropological research sometimes works: "A final report from the Office of the Ethnologist! We have at last tracked down (materialized?) an isi Vivane!... We climbed halfway up the mountain behind (Clarkebury) onto a shoulder and asked at a kraal there. Never heard of it. So we toiled to the top, muttering curses, and asked at the leadman's kraal and a couple of others. Drew a blank. It then came on to rain and, as we had left the car near the bottom of the mountain, we got soaked. We decided to call it a day but half way down, on the shoulder I spoke of, we saw a large spread of small stones...(21 feet in diameter). We checked with the inmates of the first kraal 'Oh yes. That was an isi Vivane, and the place's name was esi Vivaneni. People used to spit on a stone, throw it on and say, "Ah! Sivivane!" None dares go near the place at night. I presume the thing was originally higher... It was anything but photogenic, and not frightfully impressive, being merely a very low mound of stones.

Van Warmelo continues his tale of the heap of stones, "Field researchers will be reminded of their own experiences: you stand on the very spot but nobody tells you; you have the truth under your eyes but you cannot see it, and because you don't ask the question in exactly the right way, there almost seems to be a conspiracy not to reveal it... For the purpose of the photograph...we therefore threw some of the stones that had obviously rolled down all round, back onto the heap. In view of the unanimity of early sources in referring to "heaps" I do not think this was unwarranted interference or distortion."

Miguel Fusti, "Physical Anthropology of the Canary Islands: Old and New Views," Physical Anthropology, vol. 23, no. 3, September, 1965, p. 285.

Abstract

To gain evidence for a descriptive hypothesis which required finding out how much Negro ancestry the Canary Islanders have, Fusti collected samples of the palm prints and blood types from the population. Using statistical analysis he compared the prints and the blood of the islanders to those of other groups of Caucasians and Negroes whose blood and palm prints had been analyzed by previous researchers. He concluded that Negro racial elements among the Canary Islanders is highly improbable.

Summary of Study

"The study of the anthropology of the Canary Islanders illustrates the shift in view point among human biologists during the past century from an interest in typology to the present interest in population genetics," states Fusti. For example, earlier observers who noticed the outward physical characteristics of the Canary Islanders decided that the people had a relatively high degree of Negro ancestry. But modern analysis focuses on characteristics such as palm prints and blood types which are genetically controlled (hereditary) and which make clearer distinctions between racial groups. Previous studies of the palm prints of Negroes and Caucasians have shown that the two races differ considerably on this measure. Likewise, certain elements of blood (hp'allele) appears much less frequently among Caucasians than among Negroes.

When palm prints were taken from a sample of 367 Canary Islanders, it was found that their prints were well within the range of other Caucasians prints. On a chart that listed the range of Caucasians and Negro types of palm prints, the Canary Islanders were far from the Negro types. Thus, palm prints revealed no signs of Negro influence. The Canary Islanders were also much different from Negroes in the blood types. The blood samples which were taken from 139 Canary Islanders placed the Islanders closest to Spanish blood types. Thus, again, Negroid racial elements on the island are highly improbable. Certainly, there has been far less Negro influence than other studies have indicated.

The population of the Canary Islands has changed very little since the Neolithic immigrants came from Northwest Africa. These early people were of two types: Cro-Magon and robust Mediterranean (proto-Mediterranean or Eurafricanid). Since the Canary Islands have been long isolated from each other by geographic barriers, both of the early groups of immigrants persist on the islands.

John Cobley, "The Crimes of the First Fleeters," Journal of the Royal Australian Historic Society, vol. 52, part 2, June, 1966, p. 81.

Abstract

Rather than relying on speculation and hearsay, John Cobley consulted statistical and historical records to find out just what the Australian First Fleeters had done to deserve deportation--a descriptive hypothesis.

Summary of Study

Many Australians have the doubtful distinction of being descendants of shiploads of English convicts who were deported to Australia in lieu of being hanged or otherwise more severely punished for their crimes. Two myths have arisen about the nature of the deportees, called First Fleeters. One asserts that they were villagers sentenced for trivial offenses committed out of economic necessity. The other charges that they were persistent and hardened criminals from city slums. To find out just what the First Fleeters had been tried for and where they came from, Cobley searched through the lists of convicts in Home Office papers, Old Bailey Session Papers from the London area, the Quarter Sessions Records from the County Record Offices, and the Assize records from the Public Record Office. Cobley gives no information on the ages, occupations, marital status or religions of the convicts because he says that these personal characteristics had no bearing on the charges.

Cobley uncovered the indictments of 646 of the 778 First Fleeters. There were 192 women among the convicts. Most of the charges were for stealing, burglary, robbery with violence, and rural crimes of stealing livestock. There were no poachers or political criminals in the group. Many of the women were prostitutes, who acted as decoys for robbery. Of the 646 indictments found, 190 were reprieved death sentences. About 20% of the verdicts were deduced from more serious charges. There is no way of knowing whether or not the First Fleeters were habitual criminals, but the changes indicate that there were more sinners than sinned-against villagers among the deportees.

The number of reduced sentences indicate that the judges and juries were tolerant and restrained in their use of the law. But it is certainly true that English law was harsh, inconsistent, and sometimes barbarous. English law is a baffling conglomeration of Roman law and common law (precedent). The legislature, which is composed of property owners, developed a strict code of law to protect their precious possessions. A felony was punished by death and forfeiture of property. Thus, a man could be hanged for stealing a few dollars and his family could be deprived of whatever property he had accumulated.

"The Riddle of Meda' in Salih," Archeology, vol 16. no. 3, p. 217.

Abstract

A new interpretation, based on a descriptive hypothesis, of the history of the Meda' in Salih tombs was made possible when new data from the site was collected and analyzed. Using external authentication, the historian compared the tombs and inscriptions found at Meda' in Salih with other Nabataean sites and concluded that Meda' in Salih was also a Nabataean settlement. The historian concluded that Meda' in Salih was a large settlement by the external clues of many wells and buildings and the wasp nests found in the tombs. No physical evidence of an earthquake was found. Thus, the traditional Koranic explanation of the fate of the site was discounted and another explanation was suggested.

Summary of Study

The rock-cut tombs of Meda' in Salih are on an ancient caravan route and later a pilgrimage route between the Eastern Mediterranean and Southwest Arabia. The Greeks and Romans used to get frankincense and spices from this trade route. From some time the Nabataeans, who lived South of the Dead Sea, controlled the northern part of the caravan route and established an outpost as far south as Meda' in Salih. The tombs and the inscriptions cut in their sandstone walls are indisputably Nabataean. But the tombs near al-'Ula, about 12 miles south, are just as surely not Nabataean. Because the early explorers of Meda' in Salih did not find any signs of extensive habitation in the area, it was assumed to be merely a trading post. The tombs, then, were assumed to be built for bodies which were brought from distant places.

In the past few years, the area had become accessible by car and train, and new theories about the history of Meda' in Salih must be developed to account for the new information gathered about the area. Recently Bedouin Arabs have been discovering and cleaning out old wells near the tombs. So far, they found 20 wells which lie roughly in the shape of an L several miles in length. The wells are 25 - 35 feet deep and about 15 feet in diameter. Many of them are lined with well cut sandstone blocks. The wells were dug in solid sandstone "so neatly they look as if they had been bored by a gigantic drill." Abundant water has been brought from the wells for irrigation.

The surface sherds are more extensive than previous observers have indicated. Also, the tops of walls and buildings have been unearthed in several places by the Bedouins who were searching for wells. These new finds indicate a settlement at Meda' in Salih that was much larger than the trading post previously postulated. Probably several thousand people lived in the area and carved the elaborate tombs. Also, the many abandoned wasps' nests in the tombs are a rather clear sign that the area must have been well-cultivated and thickly settled. (Wasps will live only in cultivated areas of the desert.)

The Koran says that an earthquake or a "catastrope coming from God" destroyed Meda' in Salih, but there is no evidence of an earthquake since the tombs and wells were built. Since the site was already abandoned by Muhammad's time in the 7th century, "the Koranic account may record the lingering memory of a catastrophe." It is unlikely that Meda' in Salih disappeared because of the end of the overland trade from Southern Arabia. Al-'Ula, just slightly south of the Meda' in Salih tombs, has remained a thriving area to this day. Most likely, Meda' in Salih was depopulated by either war or plague. Until recently, in fact, the tombs were avoided by the Bedouin and the whole area was considered to be accursed. Further excavations should illuminate the origin and the end of the Nabataeans.

Cyrus H. Gordon, "The Greeks and the Hebrews," Scientific American, February, 1965, p. 102.

Abstract

Gordon has collected a wide variety of data to show that the early Greek and Hebrew cultures had common roots among the Semites of the Nile Valley. He supports his conclusion by using the internal authentication techniques of finding previously overlooked similarities between ancient Greek and Hebrew literature and of showing that the language of the ancient Cretans (from the early Greeks borrowed their culture) was a Semitic language. External authentication was used to determine the Cretan-Greek origin of an ancient tomb in Syria.

Summary of Study

> The Classical civilizations of Greece and Judea have traditionally been regarded as entirely distinct cultures, yet today two lines of evidence are combining to support the hypothesis that they have a common background.

For one thing, Gordon continues, there are clear similarities between early Greek and Hebrew literature. Secondly, the oldest writing found on the island of Crete is most probably in a Semitic language.

Why have Western scholars built an artificial boundary between Greek and Hebrew cultures? Since Westerners have long doubted the historical truth of Hebrew and Greek tales, the Old Testament came to be seen only as sacred writing with no connection with historical reality, while Greek historical legends which were written about a period roughly the same as that of the Old Testament has been downgraded almost to the level of nursery tales. Thus, the image of rational Greece and religious Israel was created that gave the impression that the two cultures had no relation with each other.

In 1928 a tomb built by the people of an ancient Greek city, Mycenae, was discovered in Ras Shamra, Syria. Here was the first evidence of ancient Greek and Semitic ties. Tablets written between the 14th and 12th centuries B.C. found in the Mycenaean tomb in Syria are written in a Semitic dialect similar to two other Semitic languages, Phoenician and biblical Hebrew. The tablets hold rich epic literature which tells of relationships with Crete, an island inhabited by Greeks at an early time, and contain parallels which bridge the gap between Homer and the Bible. The local god of arts and crafts is said to have had his workshop on Crete and to have come originally from Egypt. As the Greeks and Hebrews concentrated on conquering their homelands, it is to be expected that their literatures should focus on the heroic adventures of royalty. An epic poem tells a tale strikingly similar to the story of Helen of Troy in the Iliad. In the story of Abraham and Sarah, the Bible says that they shall start a line of Kings and that Sarah was stolen from Abraham twice.

Aside from the evidence from the Mycenaean-Syrian tomb, the ancient Greeks and Hebrews wrote of their ties. The Old Testament includes Greeks in the table of nations and mentions seafaring tribes of Israel. The Greeks wrote that King Minos of Crete was the son of a Phoenician princess. The princess' brother became King of Thebes and introduced writing to Greece.

Unravelling the unknown Cretan languages has further illuminated Greek-Hebrew ties. The Minoans lived on Crete before the Greeks and the king of Minoans, King Minos had a Phoenician mother. This knowledge plus the fact that there are a number of words in the official language of King Minos' Crete and Semitic languages, especially Phoenecian clearly shows that the main language of Crete was most probably Semitic. But even more important, a 4th century Latin author, recently reassessed, wrote that the Emperor Nero recognized the ancient writings of Crete as being Phoenician. Thus, the recent deciphering of Cretan writing is really a rediscovery of what was still known in Rome in the early centuries after the birth of Christ.

> What do these text decipherments and parallels between early Greek and early Hebrew literature suggest? In my opinion, the conclusion is inescapable that both the Greek and the Hebrew heritage are rooted in a single cosmopolitan culture that flourished throughout the eastern Mediterranean during the second millennium B.C. with Crete as its major center.

Perhaps this Mediterranean, Greek-Hebrew culture developed in this way: In 4,000 B.C., before writing was widespread, Mesopotamian civilization and Egyptian civilization in the Nile Valley had risen. Both peoples traded widely in the Eastern Mediterranean area. Trade in goods and people was established throughout the area and accelerated cultural diffusion. Mesopotamia and Egypt had common commercial interests in Syria (Phoenician) and Palestine (Hebrew), where the Semitic peoples were literate.

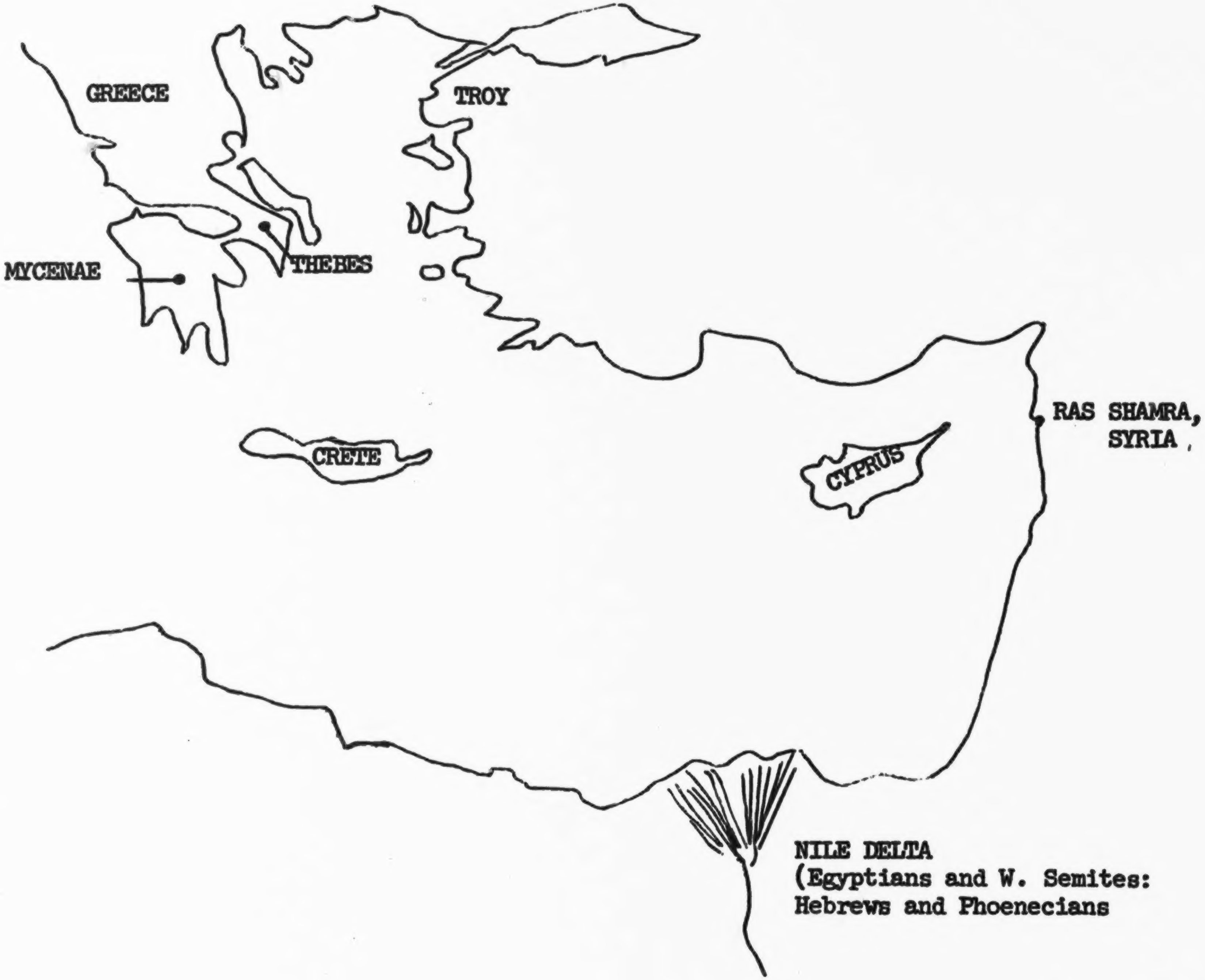
After 2,000 B.C. Indo-Europeans, who spoke languages related to Greek, Latin, etc., founded the Hittite Empire in Turkey. The Hittites were a link between "savage" Europe and "civilized" Asia and began to rival Egypt and Mesopotamia.

By 1,800 B.C. the Semitic Phoenicians were able to govern the shores of the Eastern Mediterranean. They developed Crete as their main base to control the surrounding areas. Originally, the Phoenicians of Crete were Semites from the Nile Delta. They may have been among the Semites forced out of Egypt by the 12th Dynasty Pharaohs. The Egyptian origin of the Semitic Cretes is revealed in the ancient art and architecture which is similar to that of the warm Egyptian climate.

Later, the Mycenaean Greeks brought northern style architecture such as hearths to Crete and absorbed Cretan culture into their own way of life. Since the earliest Greek civilization, Mycenean, was based on the Semitic culture of Crete, it can be said that the Semitic Cretans introduced literate culture to Europe.

Somewhat after 1,450 B.C., the Mycenaean Greeks conquered Crete and forced the Semitic Cretes to migrate to places such as Ras Shamra, Syria. Before the Hebrews conquered Palestine around 1,200 B.C., they were, like the early Cretans, Semites of the Nile Delta. Thus, the same eastern-Mediterranean culture strongly influenced both Greek and Hebrew civilizations even through the Hebrews did not leave the Nile Delta until 600 years after the Cretans.

Scholars interested in the various people of the Eastern Mediterranean should work together to find evidence from the Nile Delta which will illuminate the culture of the delta Semites who established the first high culture in Europe and the first Hebrew nation in Palestine.


GREECE
TROY
THEBES
MYCENAE
CRETE
CYPRUS
RAS SHAMRA, SYRIA
NILE DELTA (Egyptians and W. Semites: Hebrews and Phoenecians

Hooland, Lumindaine, and Sheffield, Experiments in Mass Communication, chapter 9, Yale University Press, 1949.

Abstract

In an experiment designed to test the causal hypothesis that active response and "motivation" increased learning these Yale researchers used Army recruits as subjects and film as the vehicle for presentation.

Summary of Study

In an attempt to determine the differential effects derived from the use of active participation in learning paired associates, the following procedure was employed. Two versions of a film strip on the phonetic alphabet were prepared, identical except that one required the learner's intermittent oral recitation of the phonetic equivalents of the letters of the alphabet. In order to determine the effect of motivation, half of the men in each group were told that a test would be administered following presentation of the film strip. Sampling was random, (N=742) and new recruits at an Army Reception Center were the subjects. Two types of tests were administered. Approximately 10% of the subjects were given individual oral tests on their ability to recall the phonetic names immediately following the film strip. Both accuracy and time required for the responses were recorded. The rest of the men were also immediately given a written test, which consisted of the letters of the alphabet with spaces provided for insertion of the appropriate phonetic name. Results were reported in terms of the differences between average percent of names recalled between the groups contrasted, with actual probability values reported. It was indicated that active participation increased the overall amount of learning from the film strip. When phonetic names were divided into those more easily learned and those more difficult, the inclusion of active participation is more important when material to be learned is more difficult. From Army Records, AGCT scores were derived and it was found that the less intelligent groups profited most from increased participation. It was also concluded that active participation was most effective with unmotivated, less intelligent subjects using difficult material.

Projective tests are characterized by:

1. Nonstructured Situation

   "...instead of calling for a reaction that is quite stereotyped and predictable from one person to another, it may elicit many different reactions, depending on the individuals being tested."[1]

2. Catching the Subject Off Guard

   "...entice the subject into revealing himself without his being aware of the fact that he is doing so... This is especially evident when, instead of telling about themselves, their homelife, and their reactions to their parents, children are induced to play-act with dolls dressed like grown-ups and children. At first the child arranges the furniture and people in more or less stereotyped ways,...but suddenly the observer realized there has been a change, not a quick or easily detectable one, but one that has taken several minutes and has just now reached the observer's threshold. The child is more intent on the dolls, less reactive to the observer, and he is living strangely in this new fantasy world. He begins to tell stories that have immediate parallel in his real life; he makes the family people behave as family people virtually never do...then the observer has a feeling that the blinds have gone up, and he is seeing the inner person of this child. It is as if the child were making him see this family world as the child himself see its--or, perhaps as he would like to see it."[2]

---

[1] Munn, Norman L., Psychology; the fundamentals of human adjustment, London, Harpers, 1961, p. 272.

[2] Ibid.

Thematic Apperception Test

"The Thematic Apperception Test consists of a series of 20 pictures. Each picture is ambiguous enough to permit a variety of interpretations. (In Figure 15.5 is an example of the kind of picture used, though it is not one of the test pictures.) When presented with a picture, the testee is asked to make up a story of what is happening in the picture. The story is supposed to begin with events leading up to the scene and end with an outcome. Most people, when they make up such stories, identify themselves with one of the characters in the picture, and their stories may be little more than thinly disguised autobiographies. In this way the testees may reveal feelings and desires they would otherwise hesitate to discuss openly or, in some cases, would be unwilling to admit to themselves."

"As generally used, the TAT has no standardized scoring. The tester interprets it by noting recurring themes in the stories: the characteristic needs and frustrations of the hero; the relations of the hero with members of the opposite sex, with parents, or with persons in positions of authority; and the overall emotional tone of the stories, whether depressed or overly optimistic; and so on."

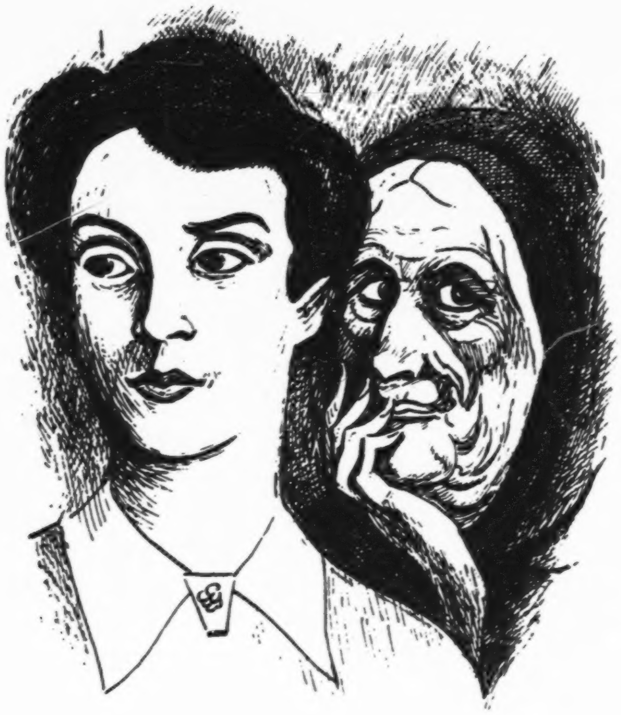
"The Rorschach consists of 10 ink blots (similar to the one in Figure 15.6, although some of the blots have colored parts to them). Each card is presented to the subject with the question, 'What might this be?' or 'What does this remind you of?' After responding to these questions for all 10 cards, the subject goes through them again, indicating what parts of the ink blot suggested his responses."

"Some of the scoring is done objectively by counting the number of times the subject responded to part of the blot compared with the number of times he responded to the blot as a whole. Counts can also be made of other things, such as the number of responses to color (in the colored blots) and the number of responses suggesting movement. On the other hand, the clinician interprets not only number of responses in different categories, but also the pattern of the responses. This becomes somewhat subjective. Even more subjective are interpretations based on other cues, such as spontaneous remarks made during the test, signs of emotional upset, and the symbolic meaning of the responses."

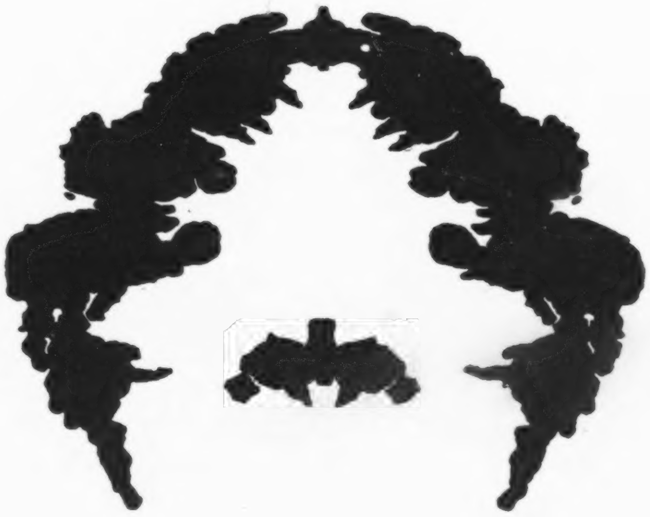
"Clinicians regularly use projective tests such as the Rorschach and TAT and believe they learn about an individual's personality from them. Perhaps they do. After all, the projective test is a subtle kind of interview. Unfortunately, convincing proof of the validity of the tests is lacking. We must therefore withhold judgment about their value as devices for personality measurement. Additional research on this subject is needed."[1]

---

[1] Morgan, Clifford T., Introduction to Psychology, McGraw-Hill, New York, 1961, p. 480.



15.5 One of the Pictures Used in the Thematic Apperception Test. (Reproduced by permission of Harvard University Press.)



15.6 An Inkblot of the Type Employed in the Rorschach Technique.

Confounding Variables

These studies of insight in animals show how much distortion can be produced by a confounding variable such as prior experience. They also illustrate that even famous psychologists can be wrong or can use inadequate methods.

Wolfgang Kohler, a pioneer in studying higher mental processes, suggested that animals can solve problems insightfully. This was a revolutionary notion which contradicted previous theories that animals solve a problem only after extensive trial and error behavior.

Kohler developed his theory of insightful learning while working with chimpanzees on an island off the coast of Africa. He placed a chimpanzee and a stick in a cage and a banana outside the cage beyond the chimp's reach. First the chimp tried in vain to reach the banana with his hand. Then, after playing with the stick for a short time, he suddenly ran to the bars of the cage and raked the banana in with the stick. It appeared that insight had occurred because the appropriate change in behavior was sudden and not after gradual trial and error.

Some of Kohler's fellow scientists scoffed at his insight theory and suggested that the study could be valid only if the prior learning of the chimp was controlled. In other words, did Kohler's chimps already know how to use sticks to reach objects that were out of reach? If so, their behavior with the stick and the banana could not be attributed to insight.

Herbert G. Birch tested Kohler's theory. He put a succession of 6 laboratory-raised chimps in a cage with a rake in the proper position for reaching the food outside the cage. Only one of the 6 chimps seemed to solve the problem by insight--and he had had previous experience in using tools. Later, Birch allowed the chimps to play freely with sticks. After playing with sticks for some time, the chimps quickly solved the food-rake problem.[1]

---

[1] Sarnoff A. Mednick, "Learning and Higher Mental Processes," Foundations of Modern Psychology Series: Learning, 1964, p. 11.

Confounding Variables--Experiment

McGuigan's amusing story of the Martain visitor indicated the vast number of confounding variables which can creep into an experiment to destroy its validity.

"In order to clarify the meaning of (confounding variables)...imagine that a visitor from Mars descends to Earth for the limited purpose of determining whether a feather, when all support is withdrawn, will fall to the ground... To accomplish this task he climbs to the top of a building and releases the feather. He then observes the feather as it floats through the air...

"He returns to Mars with his report that feathers behave peculiarly on Earth--when released, they go up. At this point his Martain brothers may ask a number of questions concerning the methods that he used in arriving at this conclusion: Is there air on Earth? If so, might not air movements have caused the feather to climb? Were there any sounds present? Maybe on Earth sounds have a peculiar effect on such objects as feathers. What about the presence of lights?...and so forth... If such extraneous variables had not been allowed to influence the feather, it might have actually fallen to the ground.

"It is the judgment of the Martian authorities that the experiment is inconclusive. Being a persistent soul, however, our Martian friend obtains a new grant from the Foundation for the advancement of knowledge of the Earth (commonly referred to as FAKE by government officials) and descends once more to Earth, this time with considerably more foresight. On arrival he sets about constructing a rather tall apparatus, air tight so that he can withdraw all the air from it, leaving a vacuum inside. Furthermore it is sound-proof, light-proof, resistant to radioactivity, etc. The Martian then climbs to the top, releases the feather, and observes that it falls to the ground. Returning to Mars he reports his results to the Martian academy of Science where his colleagues are much more receptive than before."[1]

---

[1] McGuigan, Frank J., Experimental Psychology, a methodological approach, Englewood Cliffs, N.J., Prentice-Hall, 1960, p. 100.

Experimental Design

Basic Designs*

Before-After: One group of subjects is measured both prior to and following the experimental treatment. For example, the variable is a film on the Constitution. Subjects are given a pretest on knowledge about the Constitution, then the treatment (film) is administered; finally subjects are given another test on their knowledge of the Constitution, and pretest and post-test scores are compared. Another example might be a pretest in spelling, practice with word roots (the experimental treatment) and a post-test in spelling. Improvement on the post-test could be attributed to the effect of the treatment. An advantage of this B-A design is that the difference in a particular subject's performance from pretest to post-test can be assessed.

After-Only Design: In this design no pretest is given. Instead, two groups are used. One group of subjects received the experimental treatment; the other receives either no treatment or ordinary instruction. Both groups must be given the same post-test or criterion test. The effects of the variable are assessed by comparing the performance of the experimental and the uninstructed or control group. Using the after-only design requires that the two comparison groups must be essentially equivalent. This can be assured by either 1) random assignment to experimental or control group (with a coin flip for example) 2) matching members of each group on a relevant characteristic (perhaps age, IQ score, etc.) An advantage of the after-only design is that it permits comparisons between two or more treatments.

Example: (Uninstructed control group) Half a fourth-grade class is randomly assigned to an experimental group; the other half belongs to the control group. The variable will be instruction in volleyball. The experimental group receives volleyball instruction; the control group just plays normally during recess. The next day both groups are given a test to see how well they can play volleyball and their results are compared.

Example: (Comparison between two treatments) The variable will be a type of visual display: Motion picture or 35mm slides. A class is divided so that half are randomly assigned to experimental group I and half to group II; group one sees slides about South America; group II sees a still movie about South America. The teacher narrates in both cases. Following the experimental treatments, both groups received the same test to find out who learned more about South America. Results may indicate which mode of instruction is better in this case.

When using the after-only design it is important to avoid confounding or contamination which can occur if members of the two groups (experimental and control) have an opportunity to discuss the differences in treatment.

Before-After schematic:

Groups	Test	Treatment	Test	
I	1	Yes	2	Effect of treatment demonstrated in improvement from test 1 to test 2.

After-Only design:
(Uninstructed control group)

Groups	Test	Treatment	Test	
I	No	Yes	Yes	Effect of treatment demonstrated by comparing test results of groups I and II.
II	No	No	Yes	

After-Only:
(Assessing two variables)

Groups	Test	Treatment	Test	
I	No	A	Yes	Relative merit of treatments assessed by comparing groups I and II on test.
II	No	B	Yes	

Statistical Summary Measures:
The mode, median, mean, and range

The items in a quantitative series, classified in a frequency distribution, tend to cluster about a point somewhere between the extremes, ... This tendency commonly is referred to as the central tendency of the series, and the point about which the items tend to cluster is called a measure of central tendency.

A measure of central tendency is a sort of average or typical value of the items in the series, and its function is to summarize the series in terms of this average value. It is a denominate quantity, being expressed in the same unit as the items.

There are three measures of central tendency in common use: mode, median, and the arithmetic mean. Since each of these measures or "averages" involves somewhat different techniques and interpretation, each will be treated separately in the following pages. As he learns about each, one should keep in mind that no single number can adequately describe a statistical series, and that consequently two or more series cannot be fairly compared on the basis of average values alone. In the past, educational workers in particular have tended to be preoccupied with average values to the exclusion of the other important features of their data.

THE MODE.

The mode may be defined as the item which occurs most frequently in a statistical series. When a particular type of wearing apparel, such as a suit or dress, is worn more frequently than other types during the fall season, that particular type of apparel is referred to as the mode for the fall season. If the majority of college graduates enter professional careers, the modal occupation of college graduates is professional. If we have the ten scores, 24, 22, 20, 30, 29, 22, 26, 28, 22, 25, the modal score is 22, since 22 is the score appearing most often. In the series 20, 21, 25, 24, there obviously is no modal score. The abbreviation commonly used for the mode is Mo.

USES OF THE MODE

The mode, which indicates the item of greatest frequency in a series, is an average widely used in everyday life. When newspapers use the term "on the average" they are usually referring to an outstanding or conspicuous tendency, and not to an arithmetic average. The mode is an average which is easy to understand, easy to determine, and one which best depicts the typical size of the items in a series. The fact that the mode can be determined by inspection favors its use as a rough index of the central tendency of the frequency distribution.

The mode is not affected by extreme scores, and we do not need to know the extreme scores in a series to determine the modal score. For example, if the modal salary of college presidents is desired, we need now know the highest and lowest salaries paid to college presidents to determine the modal salary for the group.

The mode of a distribution, more than other averages, is affected by changes in the grouping scheme, and is subject to wide fluctuations from one sample to the next. It is not at all reliable in small samples. The mode suffers the additional disadvantage of being incapable of algebraic treatment. For example, the mode of one distribution cannot be combined with the mode of another to determine the mode of the combined distributions. When the number of cases in a distribution is small, or in situations calling for fine grouping, the mode may have little meaning. If the modal salary of 50 workers is $3,000, for example, but if 45 of the 50 receive salaries different from $3,000, it would be absurd to report a mode.

The concept of mode as a measure of central tendency is chiefly useful because it encourages attention to bemodal and multimodal data and invites questions regarding the conditions giving rise to such data.

THE MEDIAN

A second measure of the central tendency of a statistical series is known as the median. By definition, the median is that point of the scale of scores below which one-half of the scores lie and above which one-half of the scores lie. Hence, the median, by virtue of its middle position, characterizes the central tendency of a series.

When a set of scores are ungrouped, the middle score or mid-measure ordinarily is taken as the median. If, for example, we have the five scores 8, 12, 15, 16, 19 arranged in order of size, the mid-measure is 15. If we have an even number of ungrouped scores arranged in order of size, the mid-measure is customarily defined as the point halfway between the two middle scores. Thus, in the series, 6, 8, 10, 13, 14, 16, the mid-measure is (10 + 13) + 2 or 11.5.

USES OF THE MEDIAN

The median of a distribution is the point below (or above) which one-half of the values lie.

In general, the median is easily understood and has several advantages as an average. When a series contains either a few extremely high or a few extremely low scores, relative to the majority of scores in the series, the median is perhaps the most representative average available, for it is not affected by extreme scores. When averages of such data as salaries, costs of homes, days lost by workers, and ages of people at time of marriage are needed, the median generally is to be preferred.

When the central tendency of an open-end distribution, i.e., a distribution having a bottom or top interval of unspecified length, is desired, the median is the most reliable measure that can be computed.

The median, like the mode, is a nonalgebraic measure, and medians of separate distributions cannot be combined to give the median of the combined distribution. It has the further disadvantage of being less dependable than the arithmetic mean, a point which will be discussed in the next section.

THE ARITHMETIC MEAN

In many cases when we wish to find the average of a set of scores, we simply divide the sum of the scores by the number of scores in the set. The result is popularly called the "average"; however, in statistics it is designated arithmetic mean in order to distinguish it from other averages. In discourse, the term arithmetic mean usually is shortened to mean.

The arithmetic mean is defined as the sum of the values in a series divided by the number. Using X_1, X_2, X_3, ..., X_n to represent the values of the respective N items in a series, the definition may be written

$$M = \frac{X_1 + X_2 + X_3 \ldots . + X_n}{N}$$

The definition may be stated more simply:

$$M = \frac{\Sigma X}{N} ,$$

in which X is the sum and N the number of the items in the series. The symbol Σ always refers to sum in statistics.

USES OF THE ARITHMETIC MEAN

The Arithmetic Mean is the most widely used measure of central tendency. Although usually somewhat more difficult to compute than the mode or median, its definition and meaning are easily understood. The mean perhaps best conveys the idea of average value, since it is derived from the exact values of the items in the series.

The fact that the mean is based upon the sum of the values in a series enhances its usefulness in some situations. If we have a set of independent observations of the same thing, e.g., the ratings of several judges of an individual or a set of measurements on a dimension or property of an object, the mean is extremely useful. When it can be shown, as is often the case, that the errors in a set of observations tend to be compensating, the mean of the observations is relatively unbiased.

But in other situations the fact that the mean is affected by the value of each item works against its fairness as a measure of central tendency. As was previously noted, when a series includes a few items of either high or low values, relative to the values of the majority, the mean is not a fair measure of central tendency. If, for example, the yearly incomes of six lawyers in a small town were $25,000, $6,000, $5,000 and $4,000, we would not ordinarily be satisfied with the use of the mean. The prospects of practicing law in a town in which all lawyers earned between $8,000 and $9,000 yearly would be grossly misleading.

INTERPRETATION AND USE OF MEASURES OF CENTRAL TENDENCY

When one first encounters measures of central tendency, he may become so engrossed in their calculation that he loses sight of their meaning. In the most general sense, the calculation of a measure of central tendency is a process of reducing a statistical series to a single, summarizing figure. The process is necessary in comparing and describing series for the simple reason that the mind cannot grasp the meaning of a series in all of its details.

The reduction of a series to an average value is not without danger of distorting information. Variability is an important feature of a statistical series. An average value conceals this feature, and a comparison of average values may be unfair and misleading if the series are dissimilar in variability. An average does not have meaning independent of the other characteristics of a statistical series; if fact, if a series is highly variable or irregular and rich in detail, an average may have no real meaning and serve no useful purpose at all.

APPROPRIATE USES OF AVERAGES

The question of which average to use in summarizing a given series is an important question, but one which permits no thumb-rule answers. A question which antecedes "which average" is whether any average will facilitate useful analysis and comparison.

Assuming that a given series is amenable to reduction to an average value of some sort, the selection of a particular average involves the considerations which have been dealth with in previous sections of this section.

The arithmetic mean is the most widely used and useful measure of central tendency. It is the most reliable measure, as a rule, and is simply and clearly defined. It perhaps best expresses the idea of an average value. Being an algebraic quantity, the mean is tractable in mathematical analysis. The most precise measures of variability and relationship, to be described later, involve the mean. It is generally advisable to use the mean as the measure of central tendency unless there is special reason for not using it.

The median is particularly useful in four situations. First, if a series contains a few extreme or exceptional values, the median generally gives a fairer impression of the average value of the series than the mean. It is usually the case that, when the median of a distribution is markedly different from the arithmetic mean, the former is the better average. Second, if there is doubt regarding the nature of the unit of measurement, the summation of a set of scores may be unsound. In this situation, the median as a point below and above which one-half of the scores lies is perhaps the most accurate statement of central tendency which can be made. Third, if a distribution has an upper or lower class interval of unspecified length, the median is the most reliable measure which can be obtained. Fourth, the median is a member of the percentile system, and hence is an appropriate average when a distribution is described and interpreted in terms of percentiles.

The mode is appropriate when a quick approximation to the point of concentration or "piling up" of the items in a series is desired. It is the only average available, if information regarding the value of greatest frequency or the most typical case is needed. Except for these rather unusual cases, the mode has little utility as an average in applied statistics. It is an unreliable and nonalgebraic measure. The concept of mode is primarily useful in analyzing and interpreting series having two or more points of concentration.

THE RANGE

The simplest way of describing the variability of the values in a series is to state the difference between the highest and lowest values. Such a difference is known as the range. The ranges of IQ's in Schools B, C, G, and J are:

School B, 114 - 81 - 33
School C, 131 - 74 - 57
School G, 125 - 54 - 71
School J, 104 - 64 - 40

When series are grouped, as are the IQ's,..., the individual items lose their exact values, and there is no way of determining the actual range of the series. For grouped series, either the difference between the mid-points of the highest and lowest class intervals, or the difference between the higher expressed limit in the top class and the lower expressed limit in the bottom class, may be taken as the range. Ordinarily, the range is used with reference to ungrouped series.

When we examine the distributions...we note that the ranges, although roughly indicative of dispersion or vaiability, fail to give any information about the variation of the IQ's between the extremes. The range of the great majority of IQ's in School G is from about 70 to 114; in School J from about 80 to 104. In general, the range is not a representative measure of the variability of a series. Since it is based upon the values of the two extreme items, it tells nothing about the variation of the intermediate items and is highly sensitive to sampling fluctuations.

The range is easily determined and easily understood. In a large, unimodal sample, it possesses some reliability as an estimate of variability in the population. It is chiefly useful, however, as a supplementary measure. In most situations a statement regarding range, in addition to a more representative and trustworthy measure of variability, adds to the description of the data.

External Authentication*

Here is a brief example of how external authentication can be used in anthropological research. In this case, the researchers wanted to find out whether certain stylistic details were included in the original paintings, found in a cave in Australia.

Item 2 (female anthropomorph) has 2 white double eyes; item 6 (male anthropomorph) also has eyes, but less prominently white painted. These eyes were suspect for a time. They had been chalked over. After removal of the chalk, one's suspicions remained until it was observed that white pipe-clay vertical lines were faintly present on the face, neck, torso and arms of item 2. These pipe-clay stripes are of identical substance to the pipe-clay eyes (chips examined microscopically); like the red ochre, the white is bound to the rock texture; it is therefore accepted as genuine.

Carbon-14 Dating

If a historian found a wooden coffin in an ancient Egyptian tomb, how would he know how old it is? Until recently, he would have been able to make only a rough estimate of its age. Perhaps he could guess at its age by making a thorough investigation of other artifacts in the tomb. But he would have no sure way of knowing that it was not merely a clever reproduction made by a modern Egyptian craftsman-crook.

W. F. Libbey came to the rescue of insecure scientists in 1947 when he discovered a means of dating once-living material that can usually come within a couple of hundred years of the object's age. By using radiocarbon dating of objects, scientists have discovered much about ancient men and animals as well as about changes in the earth's climate.

Radiocarbon is formed by cosmic rays--streams of atomic particles from space. As the rays reach the earth, they smash some of the atoms in the atmosphere. If a certain particle (a neutron) from a smashed atom strikes a nitrogen atom, this atom becomes radiocarbon. The radiocarbon atom immediately beging to decay at an exact and uniform rate. It loses half of itself every 5,750 years. That is only 1/2 of the radiocarbon atom is left in 5,750 years, 1/4 left after another 5,750 years, etc.

The radiocarbon content of an object can be found by burning a sample of it to convert it to carbon dioxide. After the carbon dioxide is reduced to pure carbon, the amount of radiocarbon remaining can be measured with a (radioactivity-sensitive) Geiger counter.

People, animals and plants absorb radiocarbon from the air until they die. After they die, no new radiocarbon atoms enter them. Thus, their age can be calculated rather accurately as soon as the amount of radiocarbon left in them is determined.

The Rosetta Stone as an Aid to Translation

The strange Egyptian writing that looks like abstract pictures of animals, objects and people was a mystery for centuries. No one was able to translate these hieroglyphics which adorned ancient monuments. The long-forgotten language was given up as an insolvable puzzle until 1799. In that year, a French officer of Napoleon's army discovered the Rosetta Stone half buried in the mud of the Nile River. On this amazing stone is carved a decree of Ptolemy V Epiphanu of Egypt (203-181 B.C.) in three languages--ancient Egyptian hieroglyphics, the popular language of the period, and Greek. Using the Greek inscription as a guide, a French scholar, Jean Francois Champollion labored many years to translate the hieroglyphics. By studying the position and repetition of proper names in the Greek text, he could decipher the same names in the Egyptian language. Champollion's knowledge of Coptic, the modern Egyptian language, aided his efforts. Finally, in 1822 he published his translation which solved the riddle of the ancient Egyptian language.

APPENDIX B

PRETEST

SOCIAL SCIENCE PRETEST

Version I (2-67)

NOTE: Do not mark in any way or write on this test. Please write your name on the answer sheet.

DIRECTIONS: CHOOSE THE BEST ANSWER FOR EACH OF THE FOLLOWING QUESTIONS. ON YOUR ANSWER SHEET FILL IN THE SPACE UNDER COLUMN a, b, c, OR d NEXT TO THE NUMBER OF THE QUESTION YOU ARE ANSWERING.

1. An advantage of using personal reports is that:
   a. they are rarely misinterpreted
   b. they can easily be compared statistically
   c. they are not influenced by the researcher's bias

2. A special advantage of unstructured observation is that:
   a. new hypotheses may be suggested
   b. all the observers make the same interpretation of the behavior they use
   c. information about past as well as present events can be collected

3. One characteristic of a good experiment is:
   a. the experiment is designed so it can never be repeated
   b. the situation is clearly artificial (not found in real life)
   c. the subjects do not know what variable is being studied

4. If you were an experimenter, what is the order of the steps you would take?
   a. select sample, form hypothesis, interpret results
   b. form hypothesis, interpret results, select sample
   c. select sample, interpret results, form hypothesis
   d. form hypothesis, select sample, interpret results

DIRECTIONS: DECIDE WHETHER THE FOLLOWING QUESTIONS PRESENT DESCRIPTIVE OR CAUSAL HYPOTHESES. MARK THE APPROPRIATE LETTER ON YOUR ANSWER SHEET.

5. What is the age distribution of the people living in the state of Iowa?
   a. descriptive hypothesis
   b. causal hypothesis

6. What proportion of United States citizens favors admitting Red China to the United Nations?
   a. descriptive hypothesis
   b. causal hypothesis

7. If children watch "Batman," will they become more violent?
    a. descriptive hypothesis
    b. causal hypothesis

8. If children read a lot will they have to wear glasses by the time they are 14 years old?
    a. descriptive hypothesis
    b. causal hypothesis

DIRECTIONS: READ THE FOLLOWING PARAGRAPH AND THEN CHOOSE THE BEST ANSWER FOR EACH OF THE QUESTIONS BASED ON IT.

These days there is much concern about the relationship between hard work and heart disease. A study of this relationship was made by a sociologist in London. He worked with 264 men under 65 years old who had each suffered one heart attack. For four years, one-half of the group was restricted to four hour a day work shifts. The other half of the group worked normal eight hour shifts. After six months, the blood pressure level among the men on the restricted shifts was much lower than among those on the normal shifts. However, after four years, 38 percent of those with the four hour work day and 40 percent of those with the normal work day had had second heart attacks. The sociologist concluded that shorter work shifts did not make a significant difference in the prevention of second heart attacks in the men observed.

9. This kind of hypothesis is:
    a. descriptive
    b. causal

10. The sample is:
    a. all men in London who have had one heart attack
    b. all men who work 4 hours a day
    c. 264 men

11. The condition which the researcher varies (the independent variable) in this study is:
    a. the occurrence of second heart attacks
    b. the occurrence of all heart attacks
    c. the number of hours worked

DIRECTIONS: **CHOOSE THE ONE BEST ANSWER FOR EACH OF THE FOLLOWING SHORT DESCRIPTIONS.**

12. A sociologist received a letter signed by one member of the gang which he had been studying. The letter claimed that a member of another gang shot and wounded a boy, robbed a store and slashed many automobile tires. After receiving the letter, the sociologist tried to find other samples of the letter writer's handwriting to see whether or not it matched the signature. Then he contacted him directly in order to question him about his reasons for writing the letter and about how he obtained his information.

    The procedures which the sociologist followed are examples of:
    a. internal authentication
    b. external authentication
    c. both
    d. neither

13. In an attic a man found an old photograph of an unknown relative. Because he was curious, he took the picture to a laboratory where it was determined to have been taken around the year 1920. With this information and the help of elderly relatives, the mysterious person in the photo was identified as one of his great aunts.

    These procedures are examples of:
    a. internal authentication
    b. external authentication
    c. both
    d. neither

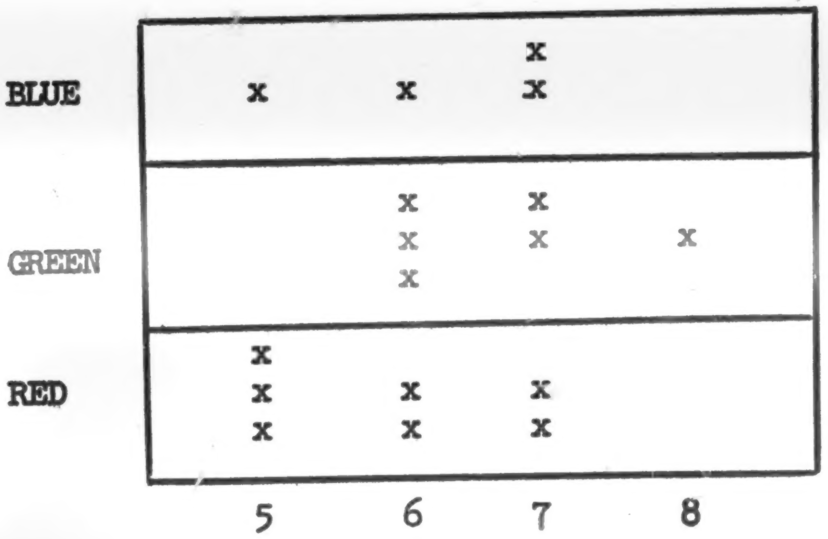
14. A social scientist came across three handwritten copies of Lincoln's Gettysburg Address. All were supposedly written in Lincoln's handwriting. He considered all three letters forgeries (not genuine) because each was written in blue ink and he had heard that Lincoln preferred black ink.

    This is an example of:
    a. internal authentication
    b. external authentication
    c. both
    d. neither

DIRECTIONS: LOOK AT THE FOLLOWING GRAPHS AND ANSWER THE QUESTION OR QUESTIONS ABOUT EACH GRAPH.

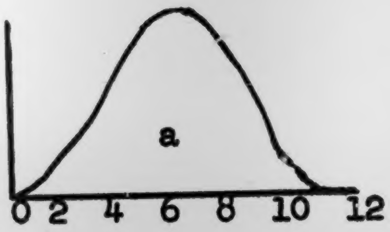
GRAPH A

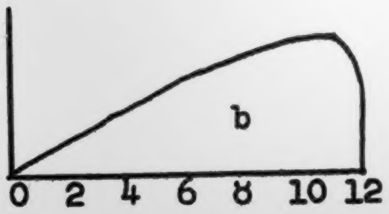
FAVORITE COLOR

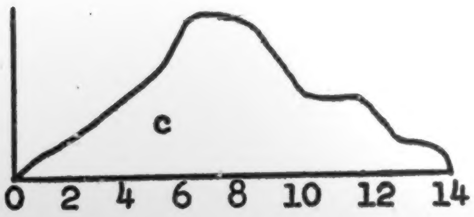


15. The range of age in Graph A is:
    a. 5 through 8 years
    b. 5, 6, 7, 8 years
    c. 3 years
    d. 4 years

16. Which of the following graphs has a median and a mean of 6 inches and a range of 12 inches?

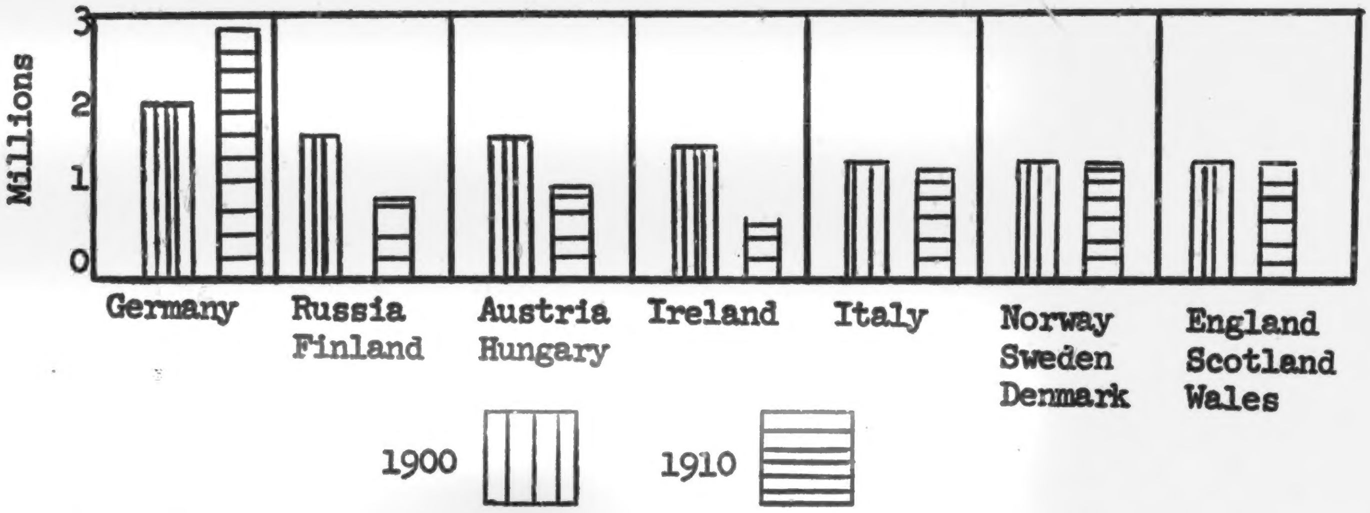






a.
b.
c.

GRAPH B
Foreign-Born Population of the United States



17. Which country contributes the most people in 1900?
    a. Germany
    b. Russia, Finland
    c. Italy

DIRECTIONS: READ THE FOLLOWING SHORT PARAGRAPHS AND CHOOSE THE SINGLE BEST ANSWER FOR THE QUESTIONS AFTER EACH.

18. A psychologist wants to find out whether students learn more in the morning or in the afternoon. He gives a spelling lesson and a spelling quiz to one eighth grade class in the morning. In the afternoon just before the students will attend the most important football game of the year, he gives the same spelling lesson and quiz to another eighth grade class. Since the afternoon class was very restless during the lesson and did much worse on the spelling quiz than the morning class, the psychologist concluded that students learn better in the morning.

    The error in research is:
    a. inadequate data for conclusion
    b. sampling error
    c. confounding variable
    d. none of the above errors

19. An advertising company is concerned about the number of people who see their ads in a particular magazine. The research staff goes to public libraries and gets the names of people who have checked out the magazine in the last six months. Then they ask these people in interviews if they have seen the ads. The results of these interviews lead them to believe the ads are effective.

The error in the research above is:
a. confounding variable
b. sampling error
c. authentication error
d. none of the above errors

20. While going through some very old letters in her attic, an historian finds a note inviting her great-grandmother to a presidential inauguration. She checks public records and finds that Grover Cleveland was inaugurated on the exact date mentioned in the note and held a fancy party following the ceremony. She concludes that her great-grandmother went to the party.

The error in the research above is:
a. sampling error
b. authentication error
c. inadequate data for conclusion
d. none of the above errors

DIRECTIONS: WHAT NECESSARY INFORMATION IS LEFT OUT OF THE FOLLOWING RESEARCH STUDIES? THERE MAY BE MORE THAN ONE ANSWER FOR EACH QUESTION.

21. Do teaching machines or regular teachers teach mathematics better? A study of this question was reported in a social science journal. One group received instruction by using red plastic teaching machines and the other group had their usual teacher. A multiple choice test consisting of twenty-five problems in area and perimeters was given to both groups and showed that students taught by regular procedures did much better.

The necessary information left out is:
a. sampling procedure
b. description of sample
c. data collection
d. statement of results
e. none of the above

22. An economist tried to find out if there is a relationship between slums and the amount of manufacturing in certain locations. He studied 18 cities which were selected at random from those with at least 250,000 people. The number of square miles of slum areas and the value of products manufactured in the cities was determined.

The necessary information left out is:
a. hypothesis
b. selection of sample
c. data analysis
d. statement of results
e. none of the above

23. In order to find out which soap consumers preferred, one hundred people were chosen at random from lists of newspaper subscribers. Their ages ranged from eighteen to fifty-five. Half were sent one brand of soap and half sent another. They were asked to comment on the soap's effectiveness in questionnaires. Brand X was judged to be best.

The necessary information left out is:
a. sampling procedure
b. description of sample
c. data collection
d. statement of results
e. none of the above

DIRECTIONS: READ THE PARAGRAPHS BELOW. THEN DECIDE WHETHER THE FOLLOWING STATEMENTS ARE a. SUPPORTED, b. CONTRADICTED, OR c. NEITHER SUPPORTED NOR CONTRADICTED BY THE FACTS GIVEN. (USE ONLY THE INFORMATION GIVEN IN EACH PARAGRAPH. DO NOT GUESS ON THE BASIS OF YOUR OWN KNOWLEDGE OF THE SUBJECT.)

An anthropologist lived in a small village in Ceylon, an island off the southern tip of India. By observation and interview he gathered facts about the marriage customs of the Veddas, one of the groups of people who live on the island. The correct marriage among the Veddas of Ceylon is for a man to marry his father's sister's daughter. The children of two brothers or of two sisters cannot marry, since such a marriage would be considered improper. When seeking a bride, the man goes to his future father-in-law with a present of dried deer flesh, grain, honey, or yams tied to his unstrung bow. The marriage ceremony is simple but appears to be absolutely binding since cases of divorce or separation are almost unknown. The women are jealously guarded by the men who do not allow traders or other strangers to see them.

24. Marriages between first cousins are forbidden.
a. supported by the facts given
b. contradicted by the facts given
c. neither supported nor contradicted by the facts given

25. If the children of two sisters were to marry they would be forced to leave the village.
a. supported by the facts given
b. contradicted by the facts given
c. neither supported nor contradicted by the facts given

26. In societies where there is a simple marriage ceremony, there seems to be a low divorce rate.
a. supported by the facts given
b. contradicted by the facts given
c. neither supported nor contradicted by the facts given

DIRECTIONS: CHOOSE THE SOCIAL SCIENTIST WHO FITS EACH OF THE FOLLOWING DESCRIPTIONS.

27. Through books such as Coming of Age in Samoa and Soviet Attitudes Toward Authority, this American anthropologist became famous for a psychological approach to the study of human societies.
a. Sigmund Freud
b. Margaret Mead
c. Auguste Comte
d. John Locke

28. His belief that all men have inalienable rights (rights which cannot be taken away) to life, liberty and property, influenced the American Declaration of Independence and Constitution. His most famous work, Two Treatises of Government, promoted the idea that the consent of the people is always necessary to establish and to maintain a government.
a. John Locke
b. Karl Marx
c. Herodotus
d. Plato

29. His great historical works, The French Revolution and the History of Frederick - Called the Great, dramatized his beliefs that dictatorships (one man rule) are good if they are successful and that democracies (people rule) cannot succeed unless the people are educated.
a. Thomas Carlyle
b. Adam Smith
c. Ivan Pavlov
d. Alexander von Humboldt

DIRECTIONS: CHOOSE THE ANSWER WHICH BEST COMPLETES THE FOLLOWING DEFINITIONS.

30. The score in the middle of a distribution or series of scores is:
a. median
b. mode
c. range

31. The difference between the highest and lowest of a group of scores is called the:
    a. variable
    b. range
    c. average

32. Generalizable means most nearly:
    a. capable of being established as truth
    b. capable of being broadly stated
    c. capable of applying to the entire group

33. A set of measurements arranged in order from lowest to highest.
    a. graph
    b. distribution
    c. range

APPENDIX C

POST-TEST

SOCIAL SCIENCE POST-TEST

Version I (2-67)

NOTE: Do not mark in any way or write on this test booklet. Please write your name on the answer sheet.

DIRECTIONS: CHOOSE THE BEST ANSWER FOR EACH OF THE FOLLOWING QUESTIONS. ON YOUR ANSWER SHEET FILL IN THE SPACE UNDER COLUMNS a, b, c, or d NEXT TO THE NUMBER OF THE QUESTION YOU ARE ANSWERING.

1. An advantage of using personal reports is that:
   a. they are rarely misinterpreted
   b. they can easily be compared statistically
   c. they are not influenced by the researcher's bias

2. An advantage of questionnaires in research is that:
   a. they can be used with people regardless of educational level
   b. they reduce the influence of the researcher on the subjects' answers
   c. they demonstrate cause and effect relationships

3. An advantage of projective tests in research is that:
   a. there is only a slight chance of misinterpreting the results
   b. they can uncover hidden attitudes and opinions
   c. they are easy to develop by researchers

4. A special advantage of unstructured observation is that:
   a. new hypotheses may be suggested.
   b. all the observers make the same interpretation of the behavior they see
   c. information about past as well as present events can be collected

5. An advantage of the interview technique is that:
   a. the interviewer can influence the answers of the subject
   b. all types of people can serve as subjects
   c. the interviewer can decide not to interview subjects who have no opinions about the research being done

6. One limitation in using the observation method to collect data is:
   a. the presence of the observer may change the behavior of the individuals
   b. the observations are not usually made by trained researchers for a research study
   c. a spoken response is required of the observed subjects

7. An experiment planned to establish a cause-effect relationship must have:
   a. random sampling
   b. many trials for each subject
   c. a condition varied by the investigator
   d. complete statistical analysis

8. An advantage of a good experiment is that:
   a. it makes subjects feel that they are in a real-life situation
   b. it can reveal information about past as well as present societies or events
   c. it encourages subjects to be honest about their opinions and feelings
   d. it can be repeated exactly to show that identical conditions will always produce the same results

9. If you were an experimenter, what is the order of the steps you would take?
   a. select sample, form hypothesis, interpret results
   b. form hypothesis, interpret results, select sample
   c. select sample, interpret results, form hypothesis
   d. form hypothesis, select sample, interpret results

10. Choose the best order of steps for conducting an experiment.
    a. choosing plan of procedure, forming hypothesis, collecting data
    b. collecting data, choosing plan of procedure, forming hypothesis
    c. forming hypothesis, choosing plan of procedure, collecting data
    d. forming hypothesis, collecting data, choosing plan of procedure

DIRECTIONS: DECIDE WHETHER THE FOLLOWING QUESTIONS PRESENT DESCRIPTIVE OR CAUSAL HYPOTHESES. MARK THE APPROPRIATE LETTER ON YOUR ANSWER SHEET.

11. Do more people watch "Batman" or "Mr. Terrific?"
    a. descriptive hypothesis
    b. causal hypothesis

12. Does cancer in arms always require amputation?
    a. descriptive hypothesis
    b. causal hypothesis

13. What method of cooking vegetables do most city people use?
    a. descriptive hypothesis
    b. causal hypothesis

14. If children watch "Batman," Will they become more violent?
    a. descriptive hypothesis
    b. causal hypothesis

15. If children read a lot will they have to wear glasses by the time they are 14 years old?
    a. descriptive hypothesis
    b. causal hypothesis

16. Does talking to a social worker once a week help juvenile delinquents?
    a. descriptive hypothesis
    b. causal hypothesis

17. Will bright, clear maps increase students' understanding of geography?
    a. descriptive hypothesis
    b. causal hypothesis

18. Are teen-agers safer drivers than 20-30 year old adults?
    a. descriptive hypothesis
    b. causal hypothesis

19. When did the custom of throwing rice at new brides and grooms begin?
    a. descriptive hypothesis
    b. causal hypothesis

20. If teen-agers receive driving instruction in high school, will they become safer drivers?
    a. descriptive hypothesis
    b. causal hypothesis

DIRECTIONS: READ THE FOLLOWING PARAGRAPH AND THEN CHOOSE THE BEST ANSWER FOR EACH OF THE QUESTIONS BASED ON IT.

Many poeple have written letters to the newspapers or have complained to each other about the fact that women walk in the streets with rollers in their hair, or that they smoke in the streets; or that men nowadays do not stand when a lady enters the room, or that men don't wear jackets in restaurants. The Gallup Poll (a well-known organization which questions Americans on their opinions and feelings) decided to find out which practices most Americans found in poor taste. They spoke with 5,000 men and 5,000 women who lived in five randomly selected cities and fifteen randomly selected small towns in all parts of the United States. All were over 21. They asked, "Which of the following do you regard as in poor taste?" and they tabulated the answers as follows:

	Men	Women
Women wearing hair rollers in public.............	62%	65%
Women wearing short shorts in public.............	53%	67%
Men not standing when a lady enters the room.....	46%	41%
Men not wearing jackets in restaurants...........	23%	19%

On the basis of this information the researchers concluded that most Americans do think that good manners are disappearing, but that they do not agree on what good manners are.

21. The question which serves as the hypothesis of this study is:
    a. are good manners in America today disappearing?
    b. what practices do most Americans find in poor taste?
    c. how many people disapprove of women wearing rollers in public?

22. This kind of hypothesis is:
    a. descriptive
    b. causal

23. The data collection procedure used here is:
    a. questionnaire
    b. interview
    c. structured observation

24. The sample used is:
    a. all men and women over 21 living in the United States
    b. all people living in the five cities and the fifteen small towns
    c. 10,000 men and women

25. The population is:
    a. all men and women over 21 living in the United States
    b. all men and women over 21 living in the twenty places selected
    c. all men and women over 21 who have opinions about good manners

DIRECTIONS: CHOOSE THE BEST ANSWER FOR EACH OF THE FOLLOWING SHORT DESCRIPTIONS.

26. A day-by-day, first person account of a farmer's life in Kentucky was discovered in an attic. It was dated 1842. Historians said it was not an authentic (trustworthy) source of information because it included a description of President Lincoln's inauguration in 1866.

    This procedure is an example of:
    a. internal authentication
    b. external authentication
    c. both
    d. neither

27. An historian was given some letters which seemed to have been written by a soldier during the Revolutionary War. The soldier said that he thought the British troops were poorly trained and were reluctant to fight. Before he could include these letters in his book, the historian tried to find out if the letters were written during the Revolutionary War and if the soldier's statements could be correct. He first tried to find out if the soldier was listed among the men who fought the war. He also sent the letters to a laboratory to find out how old they were. Finally, he tried to learn how much the British government spent on military training and how well British soldiers were trained.

    These procedures are examples of:
    a. internal authentication
    b. external authentication
    c. both
    d. neither

28. An historian, in going through the papers of Woodrow Wilson, finds a letter from the President to his wife in which he bitterly attacks the actions of a certain Senator. Before adding this material to the biography he is writing, the historian checks the handwriting of the letter and the date to find out if Wilson really wrote the letter. Having found out that Wilson did write the letter, the historian tries to find out if Wilson was in a normal frame of mind at the time. He determines that on that day Wilson was in great physical pain, and that he was very lonely for his wife who had been away for several days. He concludes that although Wilson actually wrote the letter, his opinion about the Senator may have been only a temporary one influenced by his feelings at the time.

    The historian has used procedures for:
    a. internal authentication
    b. external authentication
    c. both
    d. neither

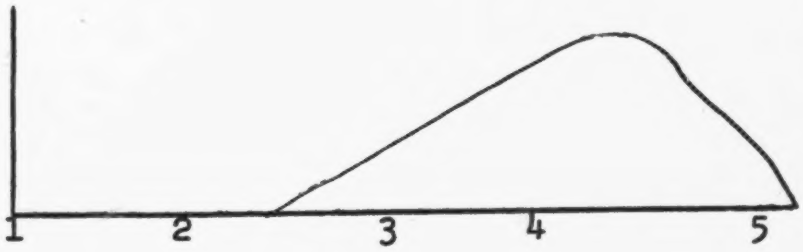
29. A researcher discovered that there were actual differences in the description of Moses crossing the Red Sea in the 16th century Gutenberg Bible and the 19th century King James Bible. He decided to accept the Gutenberg version because this Bible was older and therefore closer to biblical times.

The conclusion which the researcher reached was based on:
a. internal authentication
b. external authentication
c. both
d. neither

30. An economist, investigating the phychological impact of the 1929 stock market crash on leading business figures of the time, came across a signed and dated note among some official company papers. The note said, "I will never be the same again. This has destroyed me." At first the economist assumed that the note related to the stock market crash. However, after questioning friends of the man who had written the note, the economist discovered that on the date of the note the man's wife had just left him.

The procedure which the economist followed is an example of:
a. internal authentication
b. external authentication
c. both
d. neither

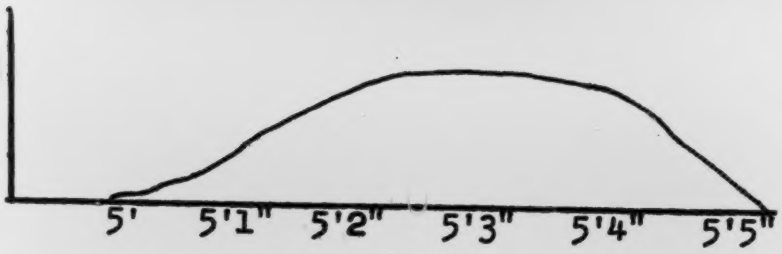
GRAPH A



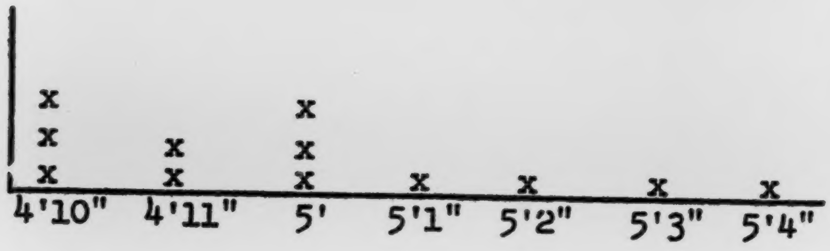
31. The mean for this distribution in Graph A is probably closest to:
a. 3
b. 4
c. 5
d. 2

32. The median of heights in a graph is five feet. The range is six inches. Which graph below has a distribution which is closest to this description?

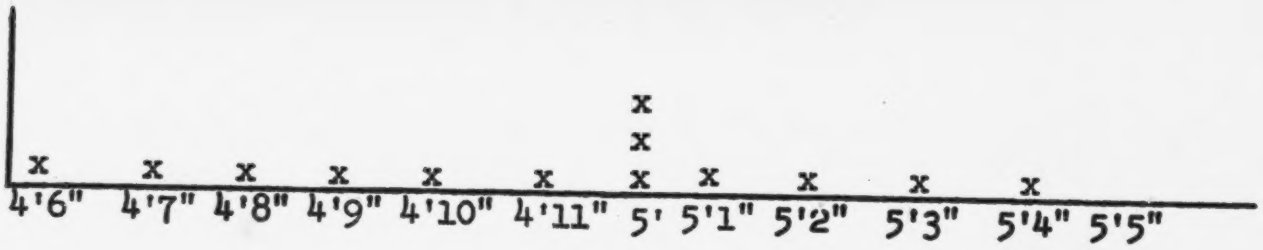
a.



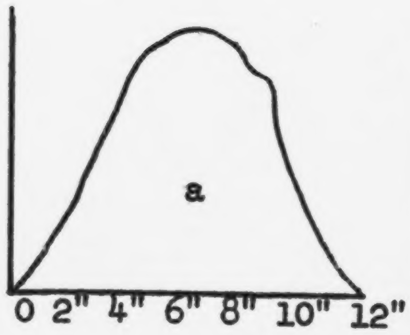
b.

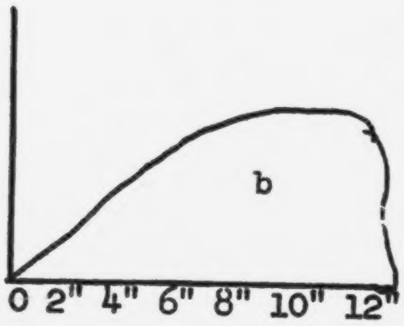


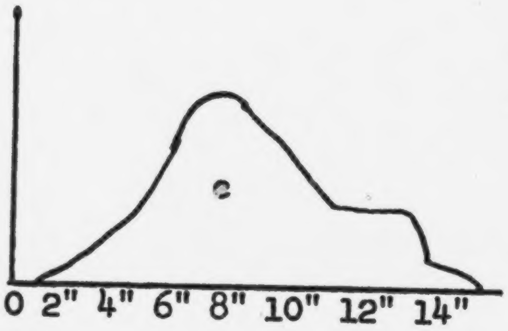
c.



33. Which of the following graphs has a median and a mean of 6 inches and a range of 12 inches?



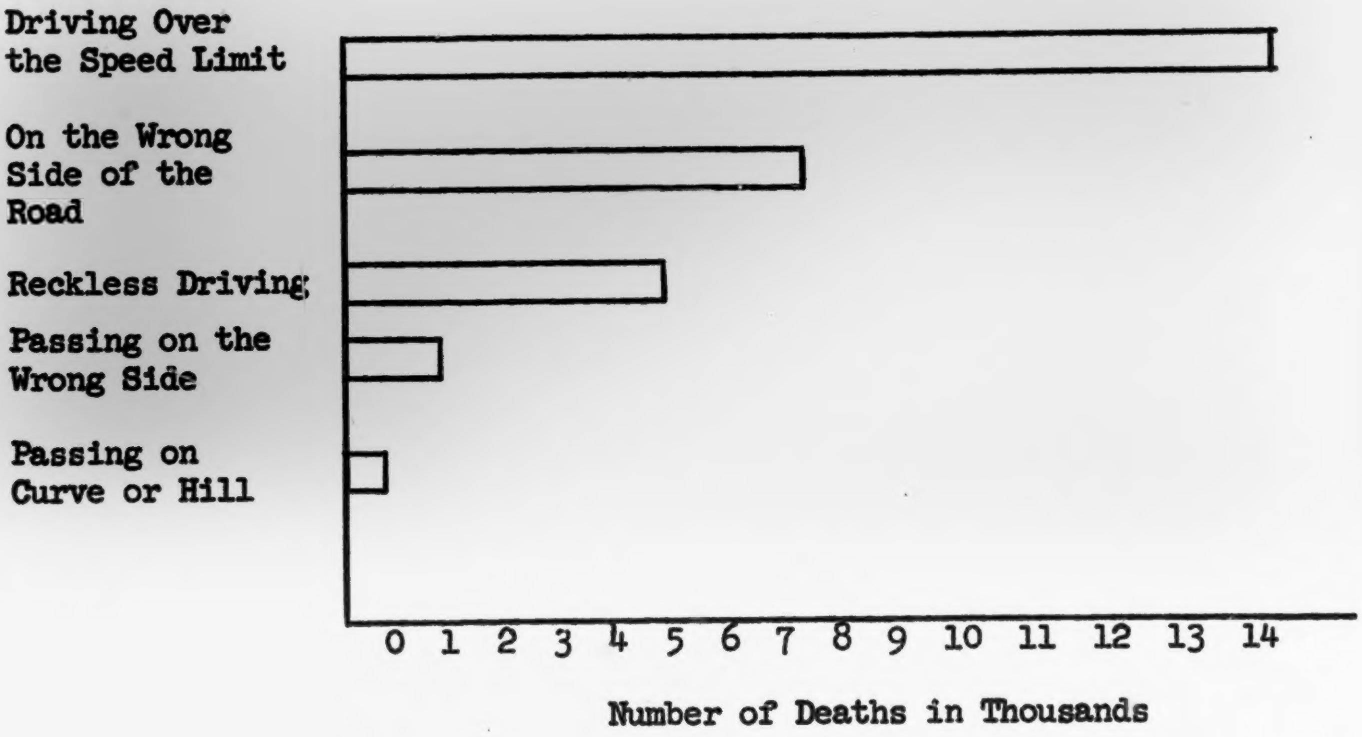




a.
b.
c.

GRAPH B

Deaths Caused by Automobile Drivers



34. Which kind of illegal automobile driving caused the greatest number of deaths in Graph B?
    a. driving over the speed limit
    b. reckless driving
    c. driving on the wrong side of the road

35. On Graph B reckless driving caused approximately:
    a. 6,000 deaths
    b. 2,500 deaths
    c. 4,500 deaths

DIRECTIONS: READ THE FOLLOWING SHORT PARAGRAPHS AND CHOOSE THE SINGLE BEST ANSWER FOR THE QUESTIONS AFTER EACH.

36. An anthropologist goes to the South Seas to find out about a primitive culture. He spends almost four years living among the people and makes many interesting discoveries about their habits. When he comes home, he writes a report describing these people.

The error in the research above is:
a. sampling error
b. confounding variable
c. inadequate data for conclusion
d. none of the above errors

37. Before he decides what colors to make his swim suits this year, a clothing manufacturer tries to find out what color his customers like. To do this, he makes a list of last year's buyers and asks them each one question, "Do you like blue?" Most of his former customers say that they do like blue. Therefore, he decides to make all of his swim suits in various shades of blue, blue-green, or blue-purple.

The error in the research above is:
a. confounding variable
b. inadequate data for conclusion
c. authentication error
d. none of the above errors

38. A team of experts is trying to determine which of three Central American countries would be the best site for a new canal to connect the Atlantic and Pacific Oceans. To find out what the people feel about a canal in their homelands, political scientists give a short questionnaire about the proposed canal to every tenth person listed in the last census in each of the three countries. After analyzing thousands of these questionnaires, the political scientists find that the people of one of these countries like the idea of a canal far better than the people of the other countries.

The error in the research above is:
a. sampling error
b. confounding variable
c. lack of authentication
d. none of the above errors

39. A teacher wanted to know if eating a large lunch interfered with her class's attention. For eight days, she supplied delicious fried chicken, turkey, roast beef sandwiches and ice cream for one half of her 5th grade class, alternating so one one day one half of her class got the treat and on the next day the other half was fed well. She discovered that the half of her class who had the large lunch seemed to pay less attention in class than the other half who ate in the cafeteria.

The error in the research above is:
a. confounding variable
b. sampling error
c. inadequate data for conclusion
d. none of the above errors

40. A hair dye manufacturer wants to know if the Mid-west will be a profitable area for his goods. He goes to a senior citizens' home in Kansas and offers free hair coloring to all ladies who wish it. Everyone volunteers. The manufacturer decides to open a warehouse in the immediate area.

The error in the above research is:
a. authentication error
b. sampling error
c. confounding variable
d. none of the above errors

DIRECTIONS: WHAT NECESSARY INFORMATION IS LEFT OUT OF THE FOLLOWING RESEARCH STUDIES? THERE MAY BE MORE THAN ONE ANSWER FOR EACH QUESTION.

41. A study was reported in which the question "Who is smarter - boys or girls?" was investigated. Four hundred ninth grade students were selected at random and their grades in English, history, mathematics and science were compared. The researchers found that boys and girls tended to earn about the same number of high grades, but that girls showed a greater superiority in English and history while boys excelled in mathematics and science.

The necessary information left out is:
a. sampling procedure
b. data collection
c. data analysis (by authentication)
d. statement of results
e. none of the above

42. The importance of modern transportation was reported by a group of government researchers. Their study showed that ten out of eleven products found in the home were transported by truck, airplane, ship or train for at least fifty miles. Such results led them to conclude that a strike including all major transportation types would seriously affect the lives of everyone in the United States.

The necessary information left out is:
a. selection of sample
b. data collection
c. statement of results
d. interpretation of results
e. none of the above

43. A sociologist reported his attempt to find a relationship between popularity of teachers and the leniency of their grading systems. He asked students to comment on questionnaires about their favorite teachers and then obtained a list of grades these teachers have given during the last two years. Much to his surprise, he found that teachers who gave lower grades were often the most popular and concluded that students like teachers who make them work hand.

The necessary information left out is:
a. hypothesis
b. sampling procedure
c. data analysis
d. interpretation of results
e. none of the above

DIRECTIONS: READ THE PARAGRAPHS BELOW. THEN DECIDE WHETHER THE FOLLOWING STATEMENTS ARE a, SUPPORTED, b, CONTRADICTED, OR c, NEITHER SUPPORTED NOR CONTRADICTED BY THE FACTS GIVEN. (USE ONLY THE INFORMATION GIVEN IN EACH PARAGRAPH. DO NOT GUESS ON THE BASIS OF YOUR OWN KNOWLEDGE OF THE SUBJECT.)

After studying letters, histories, government documents, and man-made objects from ancient Rome, an historian gathered the following information: a poor free laborer needed about fifteen bushels of wheat a year. About the only meat he ate was that which the priests gave away after a sacrifice on holidays. Each day he needed only a penny's worth of oil, a penny's worth of wine, and another penny's worth of vegetables. A pound of cheese cost a little more, but it lasted several days. As for clothing he spend about one dollar for the wool for the two tunics he needed each year and about 50¢ for the sandals he seldom wore. The state provided him free amusements on holidays and free public baths, where friends gathered. If he were out of work, the state would give him free grain. Therefore, it was possible for a poor freeman to live and even support a wife--if she could spin and weave.

44. Even the laboring class had some recreation.
    a. supported by the facts given
    b. contradicted by the facts given
    c. neither supported nor contradicted by the facts given

45. The poor freeman ate mostly cheese and vegetables since he did not like meat.
    a. supported by the facts given
    b. contradicted by the facts given
    c. neither supported nor contradicted by the facts given

46. The laborer's wife often made his clothing for him.
    a. supported by the facts given
    b. contradicted by the facts given
    c. neither supported nor contradicted by the facts given

47. Romans frequently went barefoot.
    a. supported by the facts given
    b. contradicted by the facts given
    c. neither supported nor contradicted by the facts given

48. The wheat and amusement which the Romans supplied to the freeman was one of the causes of the decline of the Roman empire.
    a. supported by the facts given
    b. contradicted by the facts given
    c. neither supported nor contradicted by the facts given

DIRECTIONS: CHOOSE THE SOCIAL SCIENTIST WHO FITS EACH OF THE FOLLOWING DESCRIPTIONS.

49. This psychologist taught dogs (by conditioned learning) to respond to the sound of a bell in the same way that they usually respond to food. These experiments showed that the mind and body are not separate but that the mind can be studied by scientific methods just as the body can.

    a. Sigmund Freud
    b. Ivan Pavlov
    c. Margaret Mead
    d. Alexander Von Humboldt

50. As a pioneer of psychoanalysis, he originated the method of free association to study the normal and abnormal mind. He revealed the influence of the unconscious mind on men's actions.

    a. Herodotus
    b. Ivan Pavlov
    c. Sigmund Freud
    d. Bronislaw Malinowski

51. He gave sociology its name and established this social science by classifying and describing the social thought which preceded him. In his attempt to use sociology as a tool for social reform, he relied on careful observation and comparison of societies.

    a. Auguste Comte
    b. John Locke
    c. Alexander Von Humboldt
    d. Thomas Carlyle

52. This famed economist was the father of the classical theory of economics. He argued in his book, Wealth of Nations, that a nation's progress and wealth could best be achieved by free enterprise, that is, without government interference in the business or trade.

    a. Karl Marx
    b. Adam Smith
    c. Plato
    d. Ivan Pavlov

53. His book, Capital, revealed his theory that economics determine the course of history. He argued that flaws in the capitalist system would certainly cause the downfall of that period and pave the way for the next stage, communism.

    a. Karl Marx
    b. John Locke
    c. Thomas Carlyle
    d. Adam Smith

54. His Republic is considered to be the first work of political science. He introduced the idea that governments should be planned by thinking men, not accepted by faith.

a. Bronislaw Malinowski
b. Adam Smith
c. Plato
d. Herodotus

55. Known as the father of history, his great work, Histories, is a masterpriece of literature and careful historical research into the background and the battles of the Greek-Persian War of 490 B.C.

a. Plato
b. Adam Smith
c. Herodotus
d. Thomas Carlyle

56. A pioneer of scientific geography, he travelled to different parts of the world in order to directly observe the geographic conditions which he classified and described.

a. Alexander Von Humboldt
b. Plato
c. Bronislaw Malinowski
d. Sigmund Freud

57. His great historical works, The French Revolution and the History of Frederick, Called the Great, dramatize his beliefs that dictatorships (one man rule) are good if they are successful and that democracies (people rule) cannot succeed unless the people are educated.

a. Thomas Carlyle
b. Adam Smith
c. Ivan Pavlov
d. Alexander Von Humboldt

58. Through his books, such as The Family Among Australian Aborigines and A Scientific Theory of Culture, this famous anthropologist developed his idea that the social institutions (family, religious organization, etc.) of primitive people are not ridiculous or wrong but are merely different ways of serving basic human physical and social needs.

a. Thomas Carlyle
b. Bronislaw Malinowski
c. Auguste Comte
d. Ivan Pavlov

DIRECTIONS: CHOOSE THE ANSWER WHICH BEST COMPLETES THE FOLLOWING DEFINITIONS.

59. The most popular or frequently appearing score is:
    a. median
    b. mode
    c. range

60. The difference between the highest and lowest of a group of scores is called the:
    a. variable
    b. range
    c. average

61. The systematic noticing and recording of information necessary to the research purpose is:
    a. interview
    b. personal reports
    c. observation

62. A carefully planned and controlled research study in which a certain condition is varied in order to find out if the condition causes certain effects:
    a. experiment
    b. investigation
    c. analysis

63. Variable means most nearly:
    a. something which can change certain aspects of its character without losing its basic identity
    b. the difference between the highest and lowest of a group of scores
    c. something which cannot change at all without losing its basic identity

64. An approximate method of determining the age of objects which are made out of once-living material.
    a. internal authentication
    b. statistical analysis
    c. carbon-14 dating

65. The study of the production, distribution and sharing of resources:
    a. history
    b. geography
    c. economics

66. An assumption which guides the collection, observation, and interpretation of facts and is supported or unsupported in the course of a research study.

a. hypothesis
b. conclusion
c. generalization

67. The total number of possible people, events or measurements which a researcher plans to study by means of a small portion of the whole group.

a. population
b. sample
c. conditions

68. The study of human or animal behavior, especially thought and emotions.

a. economics
b. sociology
c. psychology

APPENDIX D

STUDENT QUESTIONNAIRE

SOCIAL SCIENCE STUDENT QUESTIONNAIRE

Directions: Please help us to improve this unit by answering the following questions. Your reactions will only be used for this purpose, so be as frank as you can. Thank you.

1. Name ______________________________ Teacher's Name ____________________

2. During yesterday's class I was: (check one) _____ with my regular teacher, _____ with the visiting teacher, _____ absent.

3. How many previous senior high school courses have you had in social studies? _____

4. What is your approximate overall grade point average in senior high school (A = 4.0, B = 3.0, etc.)? _______

5. How would you rate your interest in the general field of social science? (geography, political science, history, psychology, anthropology, economics, sociology) _____ 1. high, _____ 2. above average, _____ 3. average, _____ 4. below average, _____ 5. low. (check one)

6. With respect to the topics covered in this unit, I was: (check one) _____ 1. extremely interested, _____ 2. interested, _____ 3. neutral, _____ 4. bored, _____ 5. extremely bored.

7. If more material of this type were available, I would: (check one) _____ 1. definitely study it on my own, _____ 2. probably study it on my own, _____ 3. not know whether to study it, _____ 4. probably not study it on my own, _____ 5. definitely not study it on my own.

Suppose you had to take one additional class this semester. For each pair below, choose the course you would prefer. Make a check for the course (in each pair) you select.

1. _____ Math/or/English _____
2. _____ Math/or/Science _____
3. _____ Social Studies/or/Math _____
4. _____ Art/or/Math _____
5. _____ P.E./or/Math _____
6. _____ English/or/Science _____
7. _____ English/or/Art _____
8. _____ English/or/Social Studies _____
9. _____ P.E./or/English _____
10. _____ Social Studies/or/Science _____
11. _____ Art/or/Social Studies _____
12. _____ Social Studies/or/P.E. _____
13. _____ Science/or/P.E. _____
14. _____ Science/or/Art _____
15. _____ Art/or/P.E. _____

APPENDIX E

INSTRUCTOR QUESTIONNAIRE

SOCIAL SCIENCE INSTRUCTOR QUESTIONNAIRE

Performance Test Resource Unit

NAME ________________________________ Social Security No. ___________

Mailing Address ____
street city state zip code

Directions: Please complete this questionnaire. Its primary purpose is to improve the unit, so please be candid in your responses.

The first three questions refer to the objectives in the unit. It will probably be necessary to refer to your copy of the objectives.

1. Circle the numbers of the objectives which you planned to cover prior to commencing the unit.

   1 2 3 4 5 6 7 8 9 10 11 12 13

2. Circle the numbers of the objectives which you actually covered during the instructional period devoted to the unit.

   1 2 3 4 5 6 7 8 9 10 11 12 13

3. In general, what is your overall rating of the objectives given in the unit?
   ____ excellent, _____ good, _____ average, _____ fair, _____ poor

4. How would you rate the organization of the material you received?
   _____ good, _____ fair, _____ poor

5. Did you use any of the reference materials included in the unit?_____
   If yes, how many? _____ most, _____ some, _____ a few

6. Did you use any of the learning activities suggested in the unit? __
   If yes, how many? _____ most, _____ some, _____ a few

7. Was 4 hours of instruction: _____ much too long, _____ a little too long, _____ about right, _____ a little too short, _____ too short

8. If your participation were again requested on a project in the same field but on a different topic, how would you respond? _____ definitely participate, _____ probably participate, _____ uncertain, _____ probably not participate, _____ definitely not participate.

9. What are your subjective feelings about your pupils' response to the unit?
   _____ enthusiastic, _____ interested, _____ neutral, _____ uninterested, _____ bored.

10. How did you react to the idea of directing your teaching toward objectives which were given to you? For example, did you really try to treat these objectives, or did you teach the unit pretty much the way you ordinarily would? In other words, what are your feelings about "teaching to someone else's goals?"

11. Are there any additional comments which you wish to make regarding the project? Any criticisms, suggestions, feelings, etc., will be greatly appreciated.

APPENDIX F

INVITATION AND INSTRUCTIONS FOR PARTICIPATION

INVITATION TO PARTICIPATE IN A RESEARCH PROJECT

One of the most perplexing problems associated with teacher education has been our inability to develop defensible measures of teaching proficiency which can be used to judge the relative worth of various methods of preparing teachers. Two years ago, the United States Office of Education awarded a research contract to the UCLA Department of Education to support the development of a measure of one aspect of teaching competence. This approach was to be based upon a teacher's ability to accomplish prespecified instructional objectives. While admitting that a teacher's ability to accomplish instructional objectives is not the only aspect of his competence, it was assumed that this was at least one important attribute of the teacher's ability. If the measuring procedure being developed has any merit, its purpose will be to serve as a criterion measure in the improvement of teacher education programs. We are soliciting your cooperation in this project which will be conducted this summer. A $25 honorarium has been provided in the grant for each participant.

Briefly, the test consists of a set of extremely explicit instructional objectives on the topic of Social Science Research Methods. Each of these objectives is stated in terms of the post-instruction performance expected of learners. A sample test item is included for each objective in order to make clear precisely what is intended by that objective. In addition, a collection of possible instructional activities and reference materials has been assembled. Essentially serving the role of a resource unit, these materials can, if the teacher desires, be used in attempting to accomplish the previously mentioned objectives. A participating teacher is given the objectives and resource materials approximately two weeks in advance of the time the unit is to be taught. Then, for one morning period of four hours, the teacher attempts to accomplish the objectives using any instructional procedures he chooses. In other words, the means of obtaining the objectives are totally up to the teacher. Only the ends to be accomplished are specified. At the beginning of the next day's class an examination based exclusively on the objectives is given to the teacher's students.

Although incomplete, these general goals will give you an idea of what is covered by the unit's specific objectives. For example, at the end of the unit the learner should be able to: (1) distinguish fact from opinion, (2) analyze written material critically, and (3) interpret data.

We are asking your help in determining whether this heretofore untried approach to the assessment of teacher competence has any validity. It seems that for such a test to possess a modest degree of validity it should at least be able to discriminate between experienced teachers and individuals who have no teaching experience whatsoever. In other words, it seems only reasonable that a person who has had teaching experience should be able to accomplish the objectives with more success than a person who has never taught. If this cannot be demonstrated, then there is apparently some deficiency either in our general approach or in the particular materials we have developed, for it is our a priori assumption that experienced teachers are better instructors than persons with no teaching experience. Therefore, the grant from the Office of Education requires that we try out our test materials with approximately 20 experienced teachers and 20 nonteachers. The nonteachers would be individuals, possibly housewives or college students, who possess a college background dealing with the social sciences but who have never previously taught.

The specific nature of your involvement in the project, should you be willing to participate, would be as follows: On one day, probably during the week of July 10, a nonteacher would take half of your class (randomly selected) to a separate classroom to teach the same topic. (For legal purposes, a regularly credentialed substitute teacher would remain in the rear of the nonteacher's classroom.) At the same time you would teach half your class the four hour unit. At the beginning of the next day's class you would administer an examination (30-45 minutes) to your entire class.

Performance of your students and, for that matter, of other participating teachers, would not be revealed to anyone in the district other than yourself, should you desire the information. Results would be coded immediately and used only for research purposes. Indeed, we are only interested in the group performance of the experienced teachers vs. the nonteachers. If you wish, of course, complete test results for your class would be available to you. A final report describing the total performance of both groups will also be given to each participant.

We stress again that the ultimate purpose of this research is to develop a measure which could be used as a defensible criterion for the evaluation of groups of teachers, not individual instructors. It is quite obvious that problems of dissimilar pupil abilities make it impossible to compare individual teachers. However, such variability can be compensated for in contrasts between large groups of teachers. It is the type of contrast which we hope will lead to the improvement of teacher education programs.

W. James Popham, Associate Professor
Department of Education
University of California, Los Angeles

NEEDED: PARTICIPANTS IN A RESEARCH PROJECT

Teach Four Hours--Earn Twenty-Five Dollars

One of the most perplexing problems associated with teacher education has been our inability to develop defensible measures of teaching proficiency which can be used to judge the relative worth of various methods of preparing teachers. Two years ago, the United States Office of Education awarded a research contract to the UCLA Department of Education to support the development of a measure of one aspect of teaching competence. This approach was to be based upon a teacher's ability to accomplish prespecified instructional objectives. While admitting that a teacher's ability to accomplish instructional objectives is not the only aspect of his competence, it was assumed that this was at least one important attribute of the teacher's ability. If the measuring procedure being developed has any merit, its purpose will be to serve as a criterion measure in the improvement of teacher education programs. We are attempting to recruit a number of volunteers to participate in the project which will be conducted during mid-July. A $25 honorarium has been provided in the grant for each participant.

Requirements:

-- You must be available for work during one morning (8-12), probably during the week of July 10.

-- You must have completed at least two years of college.

-- You must be a college student during the regular academic year. Preferably, you will be enrolled for college classes this summer.

-- You must have had a college major or minor in one of the following fields: history, sociology, anthropology, economics, geography, political science, or psychology.

-- You must not have had any previous teaching experience and should not have completed substantial coursework (education classes) leading to a teaching credential.

Your Participation:

The nature of each participant's involvement is the following: One morning during mid-July you would teach approximately 15 high school seniors for a full morning (8-12). The exact date is still somewhat flexible and usually we can set the date at your convenience. The teaching will take place in a regular high school and will deal with the topic "Research Methods in the Social Sciences." It is not necessary that you know anything about the topic, because complete background information, possible learning activities, etc. will be given to you 10 days prior to the time you are scheduled to teach.

In brief, you will be given instructional materials and asked to plan a teaching sequence for four hours. Then on one day in July, you will teach 10 - 15 students, using that plan. For this, you will be paid $25.

If you are interested, please call Mr. Clarence Weiler, 465-8647, or return the attached application form.

EXPERIMENTAL PROCEDURES FOR TEACHER PARTICIPANTS

IN U.S. OFFICE OF EDUCATION RESEARCH PROJECT

Thank you for agreeing to participate in the U.S.O.E. sponsored research project. Given below, as briefly as possible, are the instructions you should follow in carrying out your part of the project.

1. A member of the research project staff will contact you in order (a) to arrange the exact day on which you will teach the four hour unit and (b) to give you the name of the nonteacher who will be instructing half of your class.

2. Attached are the instructional materials needed in planning the four hour unit. Remember, you are free to teach in any way you wish, but please try to achieve the objectives given on pages 3-10 of the unit.

3. On the day preceding the date on which the teaching is to take place, have your class count off by twos ("one, two, one, two, etc."). You will teach all of the ones on the next day. The nonteacher will work with the twos. Please do not select the students for the two groups in any other way. If late comers need numbers, simply assign them a one or two at random.

4. On the day the instruction is to take place, your nonteacher will report to you before the class starts. A room for the nonteacher will have already been arranged. Send the "twos" with her (him) to that room. A regularly credentialed substitute teacher (for legal purposes) has been assigned to remain unobtrusively in the room while the nonteacher instructs. The nonteacher will be in your school only on this day.

5. Then teach the remaining students for the four hour period. You may assign homework for that evening if you wish.

6. Before class on the day after the teaching takes place, open the sealed envelope which will have been delivered to you during the previous week. Please do not open the envelope prior to that time. At the beginning of the period, distribute the tests (and answer sheets) to your entire class (now reassembled) and read the test directions enclosed in the envelope. Allow no more than 40 minutes for completion of the test. Then distribute the one-page student questionnaire and allow no more than 10 minutes for its completion. During the time the students are taking the test, please fill out the brief instructor questionnaire which is also enclosed in the envelope (you will need your copy of the instructional unit for this purpose). Place all the material in the packet and return it to the school office, where it will be picked up by our representative.

7. If you have any questions about these directions, please call either Clarence Weiler (465-8647, a San Diego number), or collect to Mrs. Lynda Scheer, at UCLA (213-478-9711, ext. 7303 or 3847).

8. Results will be sent to you as quickly as possible.

W. James Popham, Associate Professor
Department of Education
University of California, Los Angeles

INSTRUCTIONS FOR SUBSTITUTE TEACHER INVOLVED IN THE U.S.O.E. RESEARCH PROJECT

Dear Substitute Teacher:

Today will be a rather unique substitution experience for you, since you have, in essence, nothing to do! In a very restricted sense, you are participating in a United States Office of Education sponsored research project which has been approved by the San Diego City Schools. This project will be conducted on specified dates during the week of July 10 in order to test a particular research hypothesis. Non-credentialed teachers have been asked to teach a four hour unit of instruction (regarding Social Science Research Methodology) to twelfth grade students. The students are to be randomly selected from regular summer session instructors' classes so that the non-credentialed person is teaching half the group while the other students remain with their regular teacher. For legal purposes, it is necessary to have a regularly credentialed substitute teacher remain with the non-credentialed instructor.

It is important for our project, however, that you remain unobtrusively at the rear of the class and allow the instruction to be completely in the hands of the non-credentialed teacher. It is very possible that the individual in your class may have disciplinary difficulties, etc., but please allow him (or her) to cope with the problems by himself. The only requirement for your presence is a legal one and it would seriously jeopardize the results of the investigation if you were to play any kind of contributing role in the classroom instruction. We hope you understand the nature of your role during this rather unique substitution assignment, for the hypothesis being tested can be of considerable significance to education.

Sincerely,

W. James Popham
Associate Professor
Department of Education
University of California, Los Angeles

APPENDIX G

MEANS AND STANDARD DEVIATIONS PER OBJECTIVE OF TEACHERS AND NONTEACHERS

TABLE G-1: MEANS AND STANDARD DEVIATIONS PER OBJECTIVE OF TEACHER ONE AND NONTEACHER ONE

	TEACHER ONE		NONTEACHER ONE	
OBJECTIVE	$\bar{X}$	s	$\bar{X}$	s
1	7.60	2.41	7.30	2.16
2	7.50	0.84	8.60	0.69
3	9.00	2.10	9.50	1.58
4	6.60	2.11	7.60	2.45
5	6.33	0.71	7.47	1.24
6	6.90	2.33	7.60	3.06
7	6.62	2.28	7.12	1.77
8	3.20	2.14	6.60	1.89
9	7.00	4.83	7.00	4.83
10	6.00	3.16	6.00	3.16
11	9.50	1.58	10.00	0.00
12	5.80	1.75	8.00	1.63
13	4.40	2.45	5.80	2.89

TABLE G-2: MEANS AND STANDARD DEVIATIONS PER OBJECTIVE OF TEACHER TWO AND NONTEACHER TWO

	TEACHER TWO		NONTEACHER TWO	
OBJECTIVE	$\bar{X}$	s	$\bar{X}$	s
1	7.54	2.29	7.00	2.95
2	8.36	1.91	6.13	3.37
3	5.00	5.00	9.00	2.07
4	6.18	2.08	6.66	2.22
5	6.24	1.40	7.24	1.22
6	7.00	3.06	8.40	1.63
7	6.13	2.05	7.25	1.78
8	3.63	2.65	3.86	2.44
9	8.18	4.04	6.66	4.87
10	7.27	3.43	7.33	3.19
11	10.00	0.00	10.00	0.00
12	7.27	2.57	5.86	2.19
13	5.63	2.50	6.26	2.37

TABLE G-3: MEANS AND STANDARD DEVIATIONS PER OBJECTIVE OF TEACHER THREE AND NONTEACHER THREE.

	TEACHER THREE		NONTEACHER THREE	
OBJECTIVE	$\bar{X}$	s	$\bar{X}$	s
1	8.46	1.95	6.62	2.98
2	8.26	2.25	6.37	3.46
3	8.00	2.53	9.06	2.01
4	6.53	1.59	6.62	2.60
5	6.98	1.37	6.71	1.27
6	7.26	1.33	7.75	2.17
7	7.41	1.66	6.09	2.49
8	3.73	2.12	4.00	2.63
9	6.00	5.07	6.25	5.00
10	6.00	3.38	5.62	3.59
11	9.66	1.29	10.00	0.00
12	4.53	3.15	5.12	2.72
13	5.60	1.88	5.62	2.75

TABLE G-4: MEANS AND STANDARD DEVIATIONS PER OBJECTIVE OF TEACHER FOUR AND NONTEACHER FOUR.

	TEACHER FOUR		NONTEACHER FOUR	
OBJECTIVE	$\bar{X}$	s	$\bar{X}$	s
1	7.13	2.35	7.60	2.29
2	7.93	2.15	9.00	1.13
3	7.33	3.71	10.00	0.00
4	7.60	1.72	7.86	2.19
5	6.84	1.34	6.71	1.45
6	7.60	3.06	7.66	2.16
7	7.08	1.74	7.16	1.73
8	4.00	2.13	4.26	2.60
9	9.33	2.58	9.33	2.58
10	6.33	3.99	8.33	3.08
11	10.00	0.00	10.00	0.00
12	6.06	1.22	7.46	2.06
13	6.26	1.83	5.33	2.35

TABLE G-5: MEANS AND STANDARD DEVIATIONS PER OBJECTIVE OF TEACHER FIVE AND NONTEACHER FIVE.

	TEACHER FIVE		NONTEACHER FIVE	
OBJECTIVE	$\bar{X}$	s	$\bar{X}$	s
1	8.05	2.53	7.52	2.91
2	8.17	1.70	7.82	2.15
3	9.11	2.64	7.64	4.37
4	7.41	2.09	6.94	2.13
5	6.90	2.04	6.78	1.37
6	8.17	2.18	8.05	2.51
7	6.83	1.82	6.32	1.89
8	5.41	2.80	3.88	2.28
9	8.82	3.32	7.64	4.37
10	8.52	2.34	7.35	3.12
11	9.70	1.21	10.00	0.00
12	8.00	2.23	6.35	2.26
13	6.47	2.18	6.58	1.83

TABLE G-6: MEANS AND STANDARD DEVIATIONS PER OBJECTIVE OF TEACHER SIX AND NONTEACHER SIX.

	TEACHER SIX		NONTEACHER SIX	
OBJECTIVE	$\bar{X}$	s	$\bar{X}$	s
1	8.08	2.15	7.70	2.45
2	8.16	2.20	6.90	2.72
3	8.33	3.25	9.50	1.58
4	7.50	2.11	6.40	1.83
5	6.50	1.55	7.40	1.10
6	7.33	2.26	7.20	1.68
7	7.18	2.33	6.75	1.97
8	3.83	1.58	5.60	1.83
9	7.50	4.52	3.00	4.83
10	7.50	2.61	7.50	3.53
11	10.00	0.00	10.00	0.00
12	6.50	2.57	5.00	3.16
13	5.16	3.01	4.60	3.53

TABLE G-7: MEANS AND STANDARD DEVIATIONS PER OBJECTIVE OF TEACHER SEVEN AND NONTEACHER SEVEN.

	TEACHER SEVEN		NONTEACHER SEVEN	
OBJECTIVE	$\bar{X}$	s	$\bar{X}$	s
1	6.61	3.27	9.00	1.30
2	4.61	3.35	6.37	3.02
3	6.92	4.34	8.75	2.31
4	5.69	1.37	6.75	1.03
5	7.28	1.23	8.08	1.30
6	4.46	1.39	6.75	3.65
7	4.87	1.51	7.03	2.10
8	4.46	2.18	5.25	2.37
9	6.15	5.06	6.25	5.17
10	6.15	2.99	3.75	3.53
11	10.00	0.00	10.00	0.00
12	5.07	3.01	7.50	2.07
13	6.00	2.16	5.25	2.60

TABLE G-8: MEANS AND STANDARD DEVIATIONS PER OBJECTIVE OF TEACHER EIGHT AND NONTEACHER EIGHT.

	TEACHER EIGHT		NONTEACHER EIGHT	
OBJECTIVE	$\bar{X}$	s	$\bar{X}$	s
1	8.07	2.64	6.69	3.27
2	7.92	2.61	7.30	2.65
3	9.64	1.33	7.69	3.88
4	6.85	1.87	5.84	2.64
5	6.47	1.26	6.87	1.16
6	7.57	2.40	5.53	2.25
7	6.23	1.98	6.73	2.01
8	5.71	2.92	3.69	1.97
9	6.42	4.97	7.69	4.38
10	7.14	2.56	6.07	3.66
11	8.92	2.12	9.23	1.87
12	6.57	2.13	6.61	3.09
13	5.28	3.56	4.46	2.96

TABLE G-9: MEANS AND STANDARD DEVIATIONS PER OBJECTIVE OF TEACHER NINE AND NONTEACHER NINE.

	TEACHER NINE		NONTEACHER NINE	
OBJECTIVE	$\overline{X}$	s	$\overline{X}$	s
1	8.83	1.20	7.28	2.39
2	8.00	1.53	8.28	1.48
3	9.44	1.61	7.85	3.77
4	7.27	2.37	6.71	2.30
5	7.05	1.26	7.19	1.36
6	8.88	1.77	8.64	1.73
7	6.54	1.61	7.67	1.28
8	3.44	2.63	5.28	1.68
9	6.11	5.01	6.42	4.97
10	6.94	2.50	6.78	3.16
11	9.44	2.35	10.00	0.00
12	8.00	1.68	6.71	2.30
13	4.44	2.00	4.42	2.50

TABLE G-10: MEANS AND STANDARD DEVIATIONS PER OBJECTIVE OF TEACHER TEN AND NONTEACHER TEN.

	TEACHER TEN		NONTEACHER TEN	
OBJECTIVE	$\overline{X}$	s	$\overline{X}$	s
1	9.76	0.43	5.11	1.57
2	8.07	1.49	6.58	2.00
3	9.61	1.38	4.41	4.28
4	6.92	2.25	6.00	2.23
5	6.36	2.15	6.69	1.43
6	6.00	2.48	5.35	2.91
7	6.25	2.39	5.19	2.44
8	3.38	2.06	3.76	2.43
9	6.15	5.06	5.88	5.07
10	5.00	3.53	4.11	1.96
11	8.84	2.99	9.11	1.96
12	6.00	3.46	6.11	2.68
13	4.46	2.47	4.35	2.37

TABLE G-11: MEANS AND STANDARD DEVIATIONS PER OBJECTIVE OF TEACHER ELEVEN AND NONTEACHER ELEVEN.

	TEACHER ELEVEN		NONTEACHER ELEVEN	
OBJECTIVE	$\bar{X}$	s	$\bar{X}$	s
1	5.66	3.44	5.92	2.40
2	6.77	2.86	6.64	2.49
3	8.05	3.48	7.50	3.79
4	5.77	2.36	5.85	2.41
5	6.44	0.94	6.31	1.66
6	6.61	2.32	6.78	2.29
7	5.06	1.98	6.42	2.01
8	3.66	2.76	3.57	2.73
9	6.66	4.85	7.85	4.25
10	5.27	3.19	6.42	3.63
11	9.44	1.61	9.64	1.33
12	5.11	3.00	5.28	2.43
13	4.44	2.43	5.71	2.81

TABLE G-12: MEANS AND STANDARD DEVIATIONS PER OBJECTIVE OF TEACHER TWELVE AND NONTEACHER TWELVE

	TEACHER TWELVE		NONTEACHER TWELVE	
OBJECTIVE	$\bar{X}$	s	$\bar{X}$	s
1	8.12	2.75	6.20	2.78
2	7.12	2.52	7.53	1.55
3	6.87	4.03	7.66	3.71
4	7.12	1.25	6.00	2.00
5	6.50	1.32	6.19	1.43
6	7.37	2.55	6.40	2.74
7	6.25	1.51	6.33	2.03
8	3.62	2.33	3.60	1.88
9	6.25	5.00	8.00	4.14
10	6.56	3.96	5.66	3.71
11	9.06	2.01	9.66	1.29
12	6.12	2.12	6.13	2.44
13	4.37	1.50	5.86	2.77

TABLE G-13: MEANS AND STANDARD DEVIATIONS PER OBJECTIVE OF TEACHER THIRTEEN AND NONTEACHER THIRTEEN.

	TEACHER THIRTEEN		NONTEACHER THIRTEEN	
OBJECTIVE	$\bar{X}$	s	$\bar{X}$	s
1	7.66	2.09	6.62	2.82
2	7.53	2.38	6.25	3.31
3	3.33	4.08	5.62	4.03
4	7.46	1.40	5.75	2.04
5	6.44	1.17	6.87	0.86
6	6.66	2.35	5.87	2.09
7	6.25	1.56	5.15	2.13
8	4.53	2.77	3.87	2.36
9	7.33	4.57	5.00	5.16
10	4.66	4.41	6.87	3.59
11	10.00	0.00	9.68	1.25
12	6.26	1.83	5.50	2.58
13	4.53	1.92	4.50	2.12